DOCUMENT RESUME

ED 050 767 LI 002 829

AUTHOR	Vallee, Jacques
TITLE	Scientific Information Networks: A Case Study.
INSTITUTION	Stanford Univ., Calif. Computation Center.
REPORT NO	RR-2
PUB DATE	Sep 70
NOTE	72p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Astronomy, Cost Effectiveness, *Information Networks, *Information Retrieval, *Information Systems, Man Machine Systems, *On Line Systems, *Programing Languages, Systems Analysis, Time Sharing
IDENTIFIERS	DIRAC, Direct Access Project

ABSTRACT

The technical feasibility of a continental information network for astronomy has been demonstrated in the course of a two-month experiment conducted jointly by Dearborn Observatory of Northwestern University and the Stanford University Computation Center. The experiment simulated a scientific information network based on a high-level retrieval language of the non-procedural type named DIRAC. A data-base of astronomical catalogues was maintained in Palo Alto, California, and was queried remotely by a team of astronomers in Illinois. The relevant parameters of approximately one hundred time-sharing sessions were thus recorded. Analysis of the experiments in terms of operating system efficiency, user interface and cost effectiveness supports the idea that the network concept is basic to meaningful scientific documentation systems; it also indicates that generalized software is the key to cost-effective information retrieval in the environment considered and, by extension, in a variety of scientific areas that rely on a combination of bibliographic and catalogued information with a high degree of internal structure. The article reviews the problems of astronomical data structures in their relevance to language design and to the general problem of scientific information handling, and it discusses the various factors: administrative, computational and psychological, that will affect the implementation of future networks. (Author)

SCIENTIFIC INFORMATION NETWORKS: A CASE STUDY

RESEARCH REPORT NUMBER II
SEPTEMBER 1970

JACQUES VALLEE

STANFORD UNIVERSITY
COMPUTATION CENTER
INFORMATION SYSTEMS

SCIENTIFIC INFORMATION NETWORKS : A CASE STUDY

This is the second report on the research in progress in the Information Systems group at Stanford University Computation Center. The first report, entitled:

THE DIRAC LANGUAGE: CONCEPTS & FACILITIES

was released in May 1970 and can be obtained from the author. The third and final report in this Series will be entitled:

MEDICAL DATA-MANAGEMENT IN TIME-SHARING:
Findings of the DIRAC Project

and is scheduled for release in November 1970.

-:-:-:-:-:-:-:-:-:-:-:-:-:-

SCIENTIFIC INFORMATION NETWORKS : A CASE STUDY

Table of Contents

INTRODUCTION

Since the advent of time-sharing in the mid-sixties, considerable attention has been devoted to the concept of a documentation service used simultaneously by many specialists. Systems of this type have been designed, and some have been partially implemented, to explore the feasibility of this concept in such environments as library automation, medical literature retrieval and social science. Many obstacles have been identified in this field, and the results to date have been largely negative, with notable exceptions involving dedicated systems whose objectives were limited enough to be met by classical software.

There exists a class of users whose need for interactive data management has been largely ignored by the designers of these early systems: in the course of their research activities, professional scientists do not really need access to literature references as much as they need the capability to interact meaningfully with large observational records, standard tables, and private files of instrumental data. A research physicist, for instance, is a non-typical user from the point of view of current retrieval philosophies, because he is constantly consulting and operating on scientific catalogs (such as tables of standard wavelengths) but is only marginally interested in literature search, in view of the unreliability of keyword systems in a fast-changing linguistic environment, and in view of the rapid obsolescence of the information itself. Large institutions, on the other hand, are devoted to the preparation, publication and maintenance of scientific records, but they themselves do not have at their disposal an adequate computing tool of any generality. Clearly this is an aspect of data management that lends itself to automation, and yet computing facilities have so far been unable to respond creatively to these needs. Computer languages aimed at the scientific user have been designed exclusively with computation in mind, FORTRAN and ALGOL being the most common examples.

The present study describes an attempt to demonstrate the feasibility of large-scale data management in the scientific environment where literature search is a secondary rather than a primary task of the computer, viewed essentially as a research aid. This approach was based on the idea that the key to cost-effective information systems can only be found in the development of generalized software; a non-procedural language named DIRAC was therefore implemented for a limited series of experiments that included the operation for a period of two months of a network prototype with terminals in the San Francisco area and in Illinois, and using a large data-base of astronomical catalogues.

The object of the Report is to present the findings of this project and to analyze the reactions of the user community at the national and international level. The experiment was conducted in the field of astronomy, where it could be contrasted usefully with the concept of a "National Data Center" currently under consideration by specialists. However, these findings will be seen to have general validity: the transferability of the features discussed here to other scientific fields is inherent in the non-procedural concept that allows users to create, update and query files without programmer intervention.

A special objective of the study was to identify and measure for the first time the parameters affecting the performance of a generalized retrieval system in the time-sharing environment, covering three main areas of activity:

1) Catalogue preparation, data gathering, editing and updating.

2) Interactive retrieval of statistical and individual data about astronomical objects.

3) Bibliography search.

ACKNOWLEDGMENTS

Much of the work described here has been performed with the help of astronomers at Northwestern University who have contributed their time and ideas during a two-month long-distance experiment sponsored by the Stanford Computation Center. In particular, the support of Professor J.Allen Hynek, Director of Dearborn Observatory, and of Mr. Lloyd Wackerling, of the Astronomy Department, is very gratefully acknowledged. Dr.G.de Vaucouleurs, of the University of Texas, made available a tape containing his catalogue of Bright Galaxies. Messrs. Mack, Schwartz, Sargent, Shapiro, Rybski, and Dr.James Wray, have used the sample data-base which they helped in implementing, and they have offered many valuable comments reflected in the conclusions of this Report. Finally, we are greatly indebted to Mr.Roderick Fredrickson, (now with the RAND Corporation) and to the entire Systems Staff at Stanford, for making this experiment possible and exciting.

DIRAC

AND ASTRONOMICAL DATA RETRIEVAL

Dr. Jacques F. Vallee, Manager
Information Systems
Computation Center
Stanford University

Dr. J. Allen Hynek, Chairman
Department of Astronomy
Northwestern University
Evanston, Illinois

(Presented at the 1970 Convention of the Association for Computing Machinery, New York, 2 Sept. 1970. This article will appear in the Proceedings of ACM'70, to be published early in 1971)

DIRAC AND ASTRONOMICAL DATA RETRIEVAL

ABSTRACT

The technical feasibility of a continental information network for astronomy has been demonstrated in the course of a two-month exoeriment conducted jointly by Dearborn Observatory of Northwestern University and the Stanford University Computation Center. The experiment simulated a scientific information network based on a high-level retrieval language of the non-procedural type named DIRAC. A data-base of astronomical catalogues was maintained in Palo Alto, California, and was queried remotely by a team of astronomers in Illinois. The relevant parameters of approximately one hundred time-sharing sessions were thus recorded. Analysis of the experiments in terms of operating system efficiency, user interface and cost effectiveness supports the idea that the network concept is basic to meaningful scientific documentation systems; it also indicates that generalized software is the key to cost-effective information retrieval in the environment considered and, by extension, in a variety of scientific areas that rely on a combination of bibliographic and catalogued information with a high degree of internal structure.

The article reviews the problems of astronomical data structures in their relevance to language design and to the general problem of scientific information handling, and it discusses the various factors: administrative, computational and psychological, that will affect the implementation of future networks.

DIRAC AND ASTRONOMICAL DATA RETRIEVAL

Jacques F.Vallee and J.Allen Hynek

I. INTRODUCTION

This article does not present general statements about information retrieval, automatic documentation, or the potential significance of information networks to the work patterns of scientists, although it touches on these three subjects. Instead, we take the point of view of the explorer returning from a previously unknown territory with a little more technical knowledge, a greatly increased puzzlement and a considerable amount of fresh curiosity.

In other words, we have not come here to offer any predictions concerning the future impact of computers on science, but only to illustrate what happens when information-oriented software - in this case, the DIRAC language - is made available to a group of scientific researchers - professional astronomers - and is used by them to access data that previously could only be found in books, catalogues and professional journals.

This illustration will not take the form of a theoretical model or an hypothetical situation, but will be made by reference to an actual prototype implementation of an astronomical information network. The parameters of this network have been identified and measured during a six-week on-line experiment between Dearborn Observatory in Illinois and the Stanford Computation Center in California. Both the concept of scientific information networks and the analysis of the implementation in terms of computer systems will be discussed in this report. First, we should describe more closely the environment of astronomical information and show why an experiment in this limited area can have general validity, and to what extent our observations are relevant in a discussion of information systems of the next decade.

II. ASTRONOMICAL DATA STRUCTURES

The information explosion takes a peculiar form in Astronomy. An increasingly large volume of data is published every year in papers, articles, monographs and books. Among these data one finds new measurements of the parameters of stellar systems, new physical values affecting current theories and the statistical evaluations upon which they were based. Unfortunately, no global framework is available to integrate this dynamically changing information; publications are uncoordinated and reflect personal interest rather than a concerted effort to gather an homogeneous sample. Treacherous selection effects are present in all the catalogues, and integration of pertinent units from various sources is (in extreme cases) made impossible by the lack of a common standard.

All these remarks had already been made a century ago by Simon Newcomb, who wrote in his 'Reminiscences of an Astronomer':

> The work of all observing astronomers, so far as it could be used, must be combined into a single whole. But here again difficulties are met at every step. There has been, in times past, little or no concert of action among astronomers at different observatories. The astronomers of each nation, perhaps of each observatory, have gone to work in their own way, using discordant data, perhaps not always rigidly consistent, even in the data used in a single establishment.

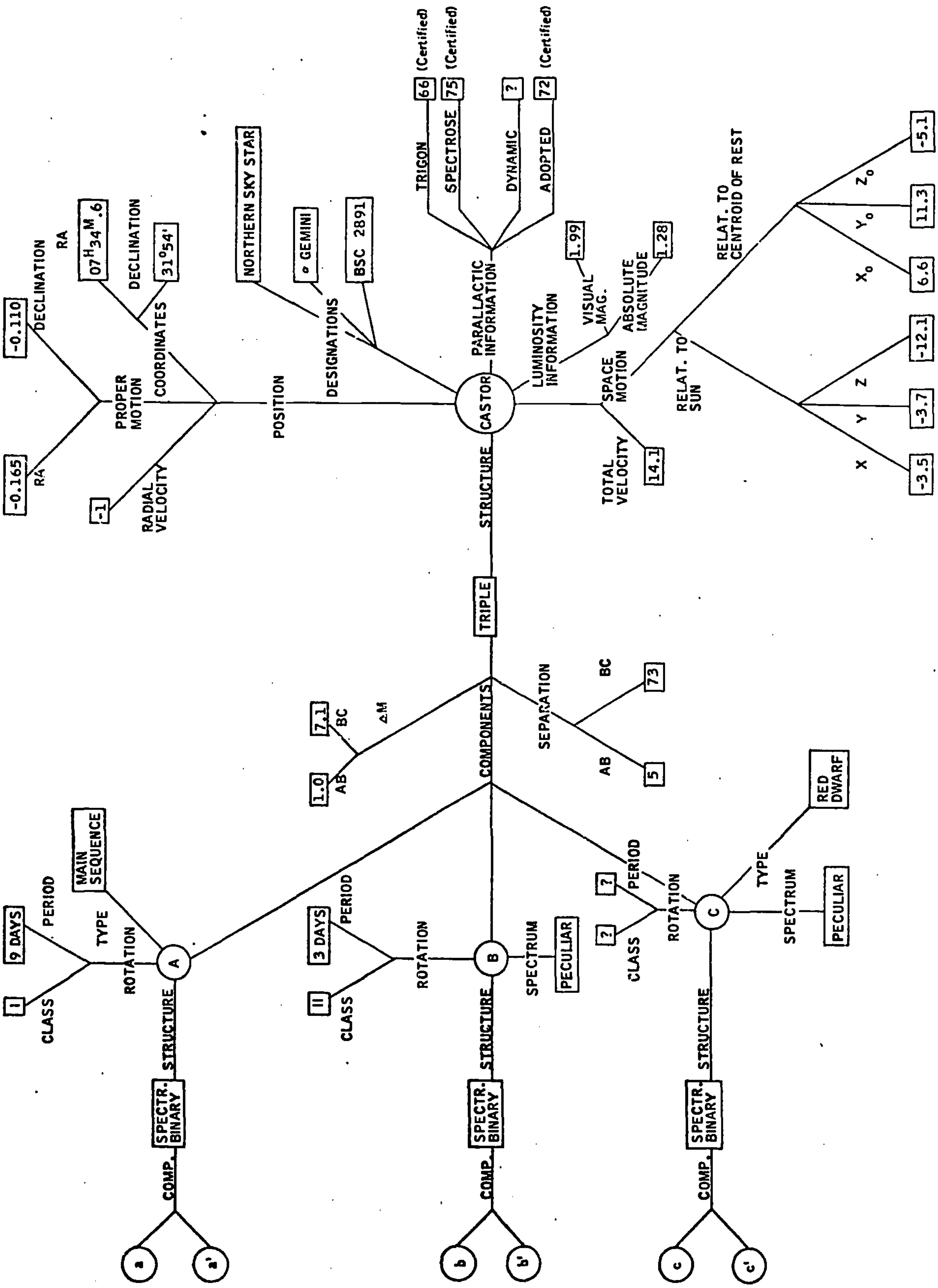
To this we must add an expanding amount of bibliographic information and general documentation. The sources of measurements for a given parameter must be clearly identified for a statistical discussion to have validity. For every adopted physical value, then, one must be able to analyze and to trace through the entire network of previous determinations - including their authors and the instrumental techniques they used. A situation is thus created which lends itself to computer analysis. At the same time, information interchange is made difficult by two things: the wide geographic dispersion of teams of astronomers; and the complexity of the information they use.

On figure 1 we have attempted to display graphically the structure of the physical information available for one stellar system, the triple star CASTOR. This naturally excludes all bibliographic information and shows only currently adopted values. It will be seen that the 'atom' of information can be explored in six main directions, each one branching into secondary levels whose structure may be a function of numerical values at higher levels. In order to adequately represent and process such structures, the designer of a retrieval system must start from the user's own terminology without forcing it a priori into a specific coding scheme. The minimum requirements for a software system supporting a file of this type would include the ability to handle large quantities of numeric information in integral form and in real form, as well as the ability to search natural-language text for keywords: this should be true at least down to the subfield level, with an unpredictable number of subfield values in any record. These requirements are not satisfied by the information-handling capabilities of procedural languages such as Fortran, Cobol or Pl-1.

III. THE DIRAC LANGUAGE.

The problem of complex information structures, that we have just mentioned, has been recognized in other fields and a number of new file organization techniques is now available to the computer scientist. The new software concept of 'non-procedural language' has emerged and is beginning to receive wide application in Business, Law, Government and military data processing (2). However, the data-base systems currently on the market, and based on this concept, have two major drawbacks when they are placed in a network environment:

Vallee/Hynek - Figure 1
Information Structure for the Star CASTOR



- These systems are generally adapted from second-generation file processing techniques and are difficult to integrate within a remote-console, time-sharing situation.
- They fail to provide a straightforward interface with text-editors and compilers in their environment.

This places upon the user a new requirement for a degree of sophistication which is simply not available in technical personnel outside computing centers. In turn, this leads to longer training periods, decreased user acceptance and considerable costs that cannot be justified in the scientific research environment.

When the idea of an astronomical network experiment was first discussed early in 1970, we were engaged in the testing of a language prototype named DIRAC (for DIRect-ACcess and also to the five types of information that it handles: Date, Integer, Real, Alphanumeric, Coded). We wanted to determine to what degree the non-procedural language concept (that is primarily business-oriented) could be extended to support information networks, in particular in the field of science and of library automation, where we felt dedicated systems had generally led to disappointing performance in spite of their high cost and the sophistication of their users.

How can we implement a truly generalized, yet cost-effective retrieval system ? The primary design objectives we proposed were a straightforward user interface, and complete integration of the language within a unified command structure in time-sharing mode. The resulting system can be best illustrated by following step by step an actual interrogation of an astronomical catalogue.

Figure 2 is an example of the on-line query of the Supernovae Catalogue implemented under DIRAC-1. The user is an astronomer who studies supernovae in the Virgo cluster. He first wants to know how many are false or suspected. The system finds one, and he displays the supernova number and the recession velocity, Vs. It will be noted that DIRAC processes information in both upper and lower case, thus simplifying the handling of textual data; this requirement is important in the applications we shall consider.

The user then wants to determine how many true supernovae in Virgo have a known Vs. The answer is 19. Restricting the search by use of the RETAIN command, he adds the rule:

$$1000 \text{ km/s} \le \text{Vs} \le 2000 \text{ km/s}$$

The answer is 11. Among these, the astronomer wants DIRAC to locate a supernova for which the first article given as reference has "Mt.Wilson" as its source. DIRAC locates supernova number 1901b. The user is now able to have the velocity, galactic coordinates, and all the literature about the object typed out on the terminal.

```
    QUERY
FILE IDENTIFICATION
:        A010
ACTION
:        SELECT
SELECTION RULES
:        Cluster CONTAINS Virgo END
______________________
24 RECORDS SELECTED
ACTION
:        RETAIN
ACTION
:        SN CONTAINS s END
______________________
 1 RECORDS SELECTED
ACTION
:        DISPLAY SN Vs Cluster
______________________
SN          s1922alpha
Vs          1243
Cluster     Virgo
______________________
 1 RECORDS SELECTED
ACTION
:        RELEASE
ACTION
:        Cluster CONTAINS Virgo AND SN DOES NOT CONTAIN s END
______________________
23 RECORDS SELECTED
ACTION
:        RETAIN
ACTION
:        Vs EXISTS END
______________________
19 RECORDS SELECTED
ACTION
:        Vs (<=2000 AND >=1000)END
______________________
11 RECORDS SELECTED
ACTION
:        Sources(FIRST) CONTAINS "Mt.Wilson" END
______________________
 1 RECORDS SELECTED
ACTION
:        DISPLAY SN Vs 12 b2 Sources END
______________________
SN          1901b
Vs          1617
12          271.15
b2          76.90
Sources   1 Ap.J.,88(1938),285-304- Contr.Mt.Wilson, 25 (1938) No.600
          2 XIV Colloque Intern.Astrophys., Paris (1941), 186, 188.
          3 Annales Observ.de Paris, 9 (1945) fasc.1, 165-179.
          4 Astronomie 55 (1941), 78, 106.
          5 Astronomie 63 (1949), 68.
          6 .............
```

Figure 2: On-line interrogation of an astronomical catalogue

IV. THE CONCEPT OF ASTRONOMICAL INFORMATION NETWORK

We now have an astronomical data-base ready, and it is accessible through a high-level language. How can such a collection of data be made available to astronomers ? What are their information and documentation requirements ? These questions can only be answered within the astronomical community itself.

With the development of large-scale numeric calculations (such as orbit computation and mathematical modeling) astronomers have become increasingly aware both of the potential usefulness and of the serious shortcomings of the computing machinery as an information processing tool. In 1969, the Astronomical Society of the Pacific expressed concern over the problems of data retrieval in astronomy and stated:

> The need for a means of recovering various data on individual stars, galaxies, clusters, or other objects is becoming obvious, and techniques for compiling, storing and distributing such data have been developed in recent years.

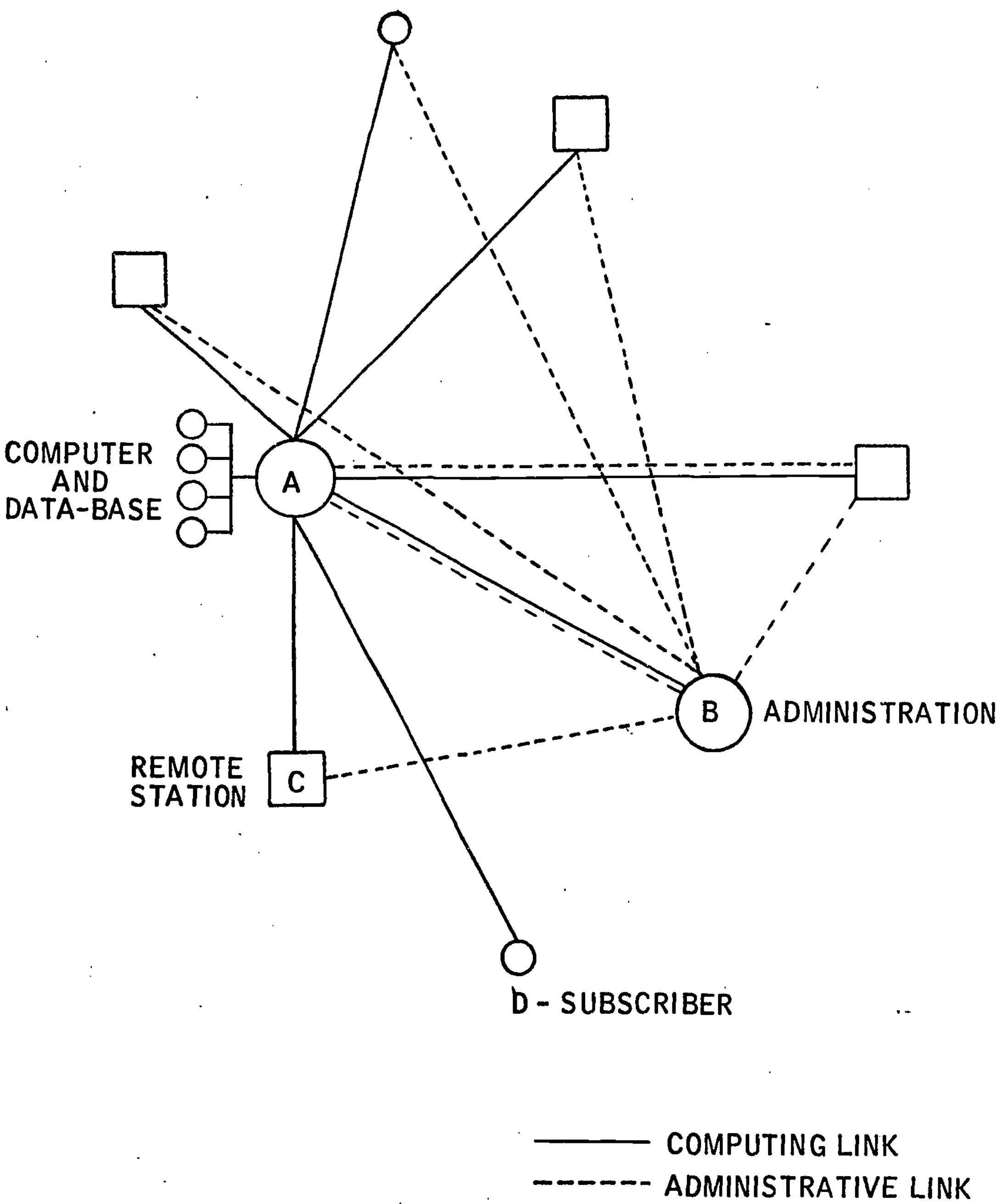
Recommending that such techniques be surveyed, the Society appointed a Committee chaired by Dr.Helmut Abt, to assess their usefulness in the astronomical environment and " to estimate the needs, procedures and cost of such a center".

In its December 1969 issue, the Society published a Special Announcement defining more exactly the role of an astronomical data center. Such a center would be

> a computerized storage, retrieval and distribution service for published data on stars, clusters, galaxies and perhaps other objects.
> Initially the catalogues available on cards or tape, both of optical and radio data, would be included: later new catalogues or bibliographies would be compiled. When practical the data itself will be listed. In other cases only bibliographies will be furnished.
> Such a center would provide two services upon request:
> 1.Listed information and/or sources for specific objects, e.g., the published data on a given star or stars, and
> 2.Derived data, e.g. the Southern F-type eclipsing stars in visual systems. (3)

While this study was conducted, Stanford and Northwestern were engaged in an experiment designed to demonstrate the feasibility of a different concept: rather than proposing as desirable the creation of a single organization (the "data center") having the responsibility to acquire and maintain the data and to implement the necessary technology, we wanted to observe the behavior of professional astronomers placed in a 'network' situation. By this we mean that our users in Illinois could not only interrogate a data-base in conversational mode but could also insert new files and update existing files in that data-base without any programmer intervention.

Vallee/Hynek - Figure 3
Scientific Information Network



According to this concept, one visualizes a network such as the one shown on Figure 3, based on a small number of basic stations equipped with remote medium-speed printers. This network would be open on a subscription basis to additional users who would receive a simple data terminal connected with a scope. The main advantages of this concept are:

1. The administration of the network (A) and the center providing computing power and storage space (B) do not have to be at the same location.

2. Responsibility for the contents of the information in storage is shared by all the users who have update privileges, but A and B do not receive special treatment.

3. Interrogation is conducted on a direct-access basis without any administrative or technical interference.

4. The network offers a highly efficient communication system among all the users of the data-base.

Although many people agree that this objective is desirable, the implementation of such a network raises technical problems from the point of view of system design. This activity has to be supported in a time-sharing environment, and several crucial parameters had never been previously measured. They included: The expected distribution of terminal sessions during the training period and during normal use of the facility; the statistical distribution of holding times; the level of interaction and growth rates of the files in the data-base. The experiment we are now going to describe was designed to measure these parameters.

V. PROTOTYPE IMPLEMENTATION OF AN ASTRONOMICAL NETWORK

An IBM 2741 communication terminal connected to Stanford's 360/67 was installed at Dearborn Observatory, North of Chicago, from April 2 to May 26, 1970. As the psychological reaction of potential users was expected to be initially negative, there was no formal announcement of the experiment and use was restricted to graduate students and staff members who had expressed interest and willingness to participate in the experiment. Previous to the installation of the terminal, several basic astronomical catalogues had been converted to machine-readable form and stored as DIRAC files at Stanford. These included the Warsaw Catalogue of Supernovae shown on figure 2, which was punched cover-to-cover for the purpose of the study; an expanded version of the Yale Bright Star Catalogue, with a volume of approximately ten million bits; and the catalogue of Bright Galaxies whose description is given on figure 4.

In eight weeks, one hundred and twenty sessions were logged, and their main parameters were recorded, leading to the time distribution of figure 5. The distribution of holding times is displayed on figure 6. Eight astronomers were involved in the experiments. Only two of them had previous computer experience at the FORTRAN level. None of them had ever been exposed to time-sharing.

Figure 5 shows (dotted line) the typical distribution of sessions on the Stanford computer, excluding administration and system development. (Average of July 1st and July 10th, when the astronomical data-base was not in use).

STANFORD UNIVERSITY
COMPUTATION CENTER

STATUS REPORT FOR FILE A070
1/MAY/1970

(CLASS 1)

LANGUAGE:
DIRAC-1A

DESCRIPTION : Catalogue of Bright Galaxies

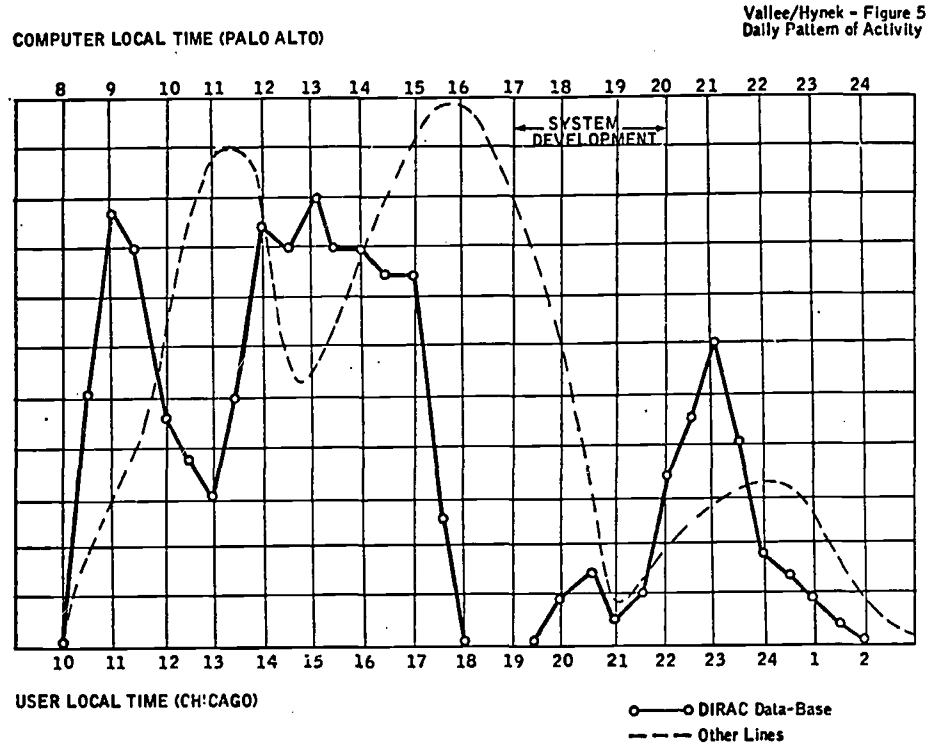
CREATION DATE : 1/MAY/1970	1 = TYPE	FILE NAME : Galaxies
RECORD NOTATION : N	2 = MULTIPLICITY	FILE CREATED BY : ~
FIELD NOTATION : @F	3 = INDEXING	RECORD LENGTH : 256
DISPOSITION : PUBLIC	4 = CODE RESIDENCE	NO.OF FIELDS : 52
NO. OF RECORDS : 2597	5 = CODE TYPE	LATEST UPDATE ON : 1/MAY/1970

F I E L D   I D E N T I F I C A T I O N ,   S T A T I S T I C S   A N D   V A L I D A T I O N   I N F O R M A T I O N

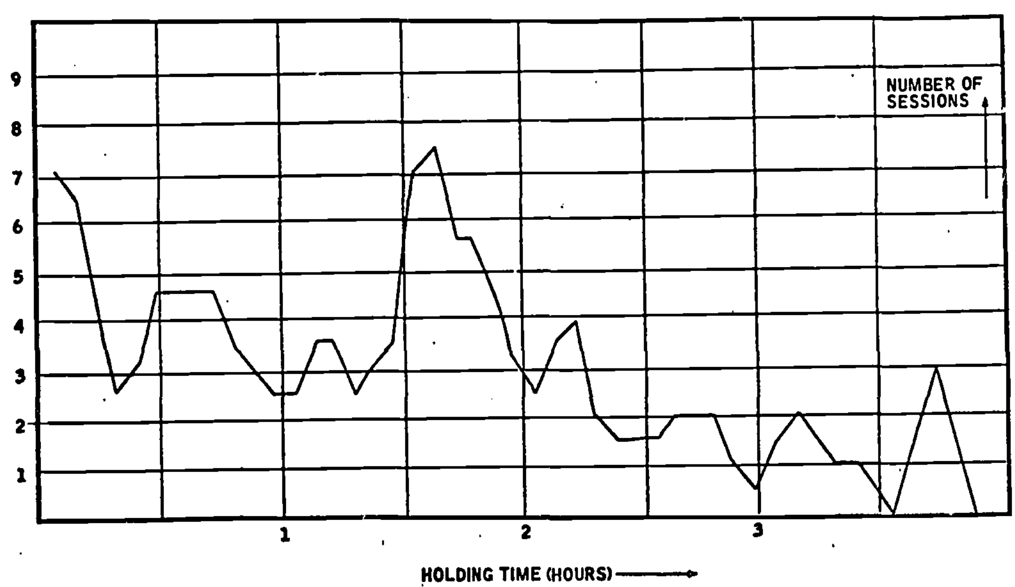
FLD	NAME	DESCRIPTION	STORAGE 1	2	3	4	5	VALIDATIONS DEC.	SIZE	SUB.	STATISTICS SIZE	DEC.	SUB.	EXISTENCE REC.	PCT
1	Name	NGC or IC number	A	S	0				6	0	0	0	0	2597	100.00%
2	HR1950	Hours of 1950 right ascension	I	S	0				2	0	0	0	0	2597	100.00%
3	MR1950	Minutes of 1950 right ascension	R	S	0				4	0	0	0	0	2597	100.00%
4	DD1950	Degrees of 1950 declination	I	S	0				3	0	0	0	0	2597	100.00%
5	MD1950	Arcminutes of 1950 declination	I	S	0				3	0	0	0	0	2597	100.00%
6	LI	Old galactic longitude, degrees	R	S	0				6	0	0	0	0	2597	100.00%
7	LII	New galactic longitude, degrees	R	S	0				6	0	0	0	0	2597	100.00%
8	SGL	Supergalactic longitude, degrees	R	S	0				5	0	0	0	0	2595	99.96%
9	dVtype	de Vaucouleurs type	A	S	0				7	0	0	0	0	2303	88.68%
10	S:dVtype	Source of de Vaucouleurs type	A	S	0				4	0	0	0	0	2189	84.25%
11	Yerkes1	Yerkes type, Morgan's list 1	A	S	0				7	0	0	0	0	609	23.45%
12	HubSan	Hubble-Sandage type	A	S	0				6	0	0	0	0	176	6.78%
13	logD	Log major diameter, 0.1 arcmin u	R	S	0				4	0	0	0	0	2293	88.29%
14	logR	Log major/minor diameters	R	S	0				5	0	0	0	0	2278	87.72%
15	logD0	Log face-on major diameter, 0.1	R	S	0				4	0	0	0	0	2278	87.72%
16	S:DR	Sources of logD and logR	A	S	0				7	0	0	0	0	2294	88.33%
17	PR100	100-yr precession, min of r.a.	R	S	0				4	0	0	0	0	2597	100.00%
18	PD100	100-yr precession, arcmin of dec	I	S	0				3	0	0	0	0	2597	100.00%
19	BI	Old galactic latitude, degrees	R	S	0				6	0	0	0	0	2597	100.00%
20	BII	New galactic latitude, degrees	R	S	0				6	0	0	0	0	2597	100.00%
21	SGB	Supergalactic latitude, degrees	R	S	0				5	0	0	0	0	2597	100.00%
22	DDO	DDO Type (Van Den Bergh)	A	S	0				10	0	0	0	0	930	35.81%
23	Yerkes2	Yerkes type, Morgan's list 2	A	S	0				7	0	0	0	0	645	24.84%
24	Holmberg	Holmberg type	A	S	0				5	0	0	0	0	300	11.55%
25	w:logD	Weight of logD	R	S	0				5	0	0	0	0	2291	88.22%
26	w:logR	Weight of logR	R	S	0				5	0	0	0	0	2275	87.64%
27	S':DR	Sources for logD and logR, as 16	A	S	0				7	0	0	0	0	170	6.55%
28	m:H	Shapley-Ames magnitude	R	S	0				4	0	0	0	0	1251	48.17%
29	n:m:H	Note regarding 28	A	S	0				1	0	0	0	0	4	0.15%
30	B0	Integrated B magnitude	R	S	0				5	0	0	0	0	873	33.62%
31	B'0	Mean B surface brightness	R	S	0				5	0	0	0	0	872	33.58%
32	S:B	Sources of B0	A	S	0				5	0	0	0	0	965	37.16%
33	B-V	Integrated blue-visual color ind	R	S	0				5	0	0	0	0	842	32.42%
34	n:B-V	Note regarding 33	A	S	0				1	0	0	0	0	87	3.35%
35	C0	Intrinsic blue-visual color inde	R	S	0				4	0	0	0	0	652	25.11%
36	S:C	Sources of color indices	A	S	0				5	0	0	0	0	[illegible]	34.55%
37	U-B	Integrated ultraviolet-blue colo	R	S	0				4	0	0	0	0	253	9.74%
38	n:U-B	Note regarding 37	A	S	0				1	0	0	0	0	28	1.08%
39	Vr	Radial velocity, km/s	I	S	0				5	0	0	0	0	958	36.89%
40	V0	Corrected radial velocity, km/s	I	S	0				5	0	0	0	0	958	36.89%
41	S:Vr	Sources of Vr	A	S	0				5	0	0	0	0	966	37.20%
42	S:photo	Sources of photographs	A	S	0				4	0	0	0	0	400	15.40%
43	m:c	Corrected Shapley-Ames magnitude	R	S	0				5	0	0	0	0	1198	46.13%
44	w:B0	Weight of B0	R	S	0				5	0	0	0	0	871	33.54%
45	m'c	Mean m:c surface brightness	R	S	0				5	0	0	0	0	1195	46.01%
46	S:B0	Sources of B0	A	S	0				5	0	0	0	0	23	0.89%
47	w:B-V	Weight of B-V	R	S	0				5	0	0	0	0	794	30.57%
48	S':C	Sources of color indices, as 36	A	S	0				5	0	0	0	0	264	10.17%
49	w:Vr	Weight of Vr	R	S	0				5	0	0	0	0	[illegible]	[illegible]
50	DeltaVr	Solar motion correction, km/s	I	S	0				5	0	0	0	0	2597	100.00%
51	S':Vr	Sources of Vr, as 41	A	S	0				5	0	0	0	0	0	0.0 %
52	S':photo	Sources of photographs, as 42	A	S	0				3	0	0	0	0	11	0.42%

Figure 4: A DIRAC Status Report: The Galaxy catalogue.

Vallee/Hynek - Figure 5
Daily Pattern of Activity



Vallee/Hynek - Figure 6
Distribution of Holding Times



17

From these figures, it is possible to make an educated guess of the expected additional load placed on a time-sharing system by a scientific documentation service of the type described, for locations with a known time difference to the computing facility. Of particular interest is the fact, reflected on figure figure 6, that sessions tend to fall into two categories: one-time interrogation series lasting from 20 to 40 minutes, and long editing and report-generating sessions with a typical duration of 90 minutes. (In comparison, the average attach time per session for all terminals connected to Stanford's 360/67, over a one-month period in late June and early July 1969 was 17.5 minutes). Use of a display scope and a remote medium-speed printer would obviously have saved much of that time. However the editing time would have remained considerable, and we do recommend the use of some local storing facility, possibly using magnetic card devices, in future systems of this type.

In terms of CPU time, we found that text-editor usage remained at a level of approximately one hour per month while DIRAC execution and file system access required three times as much. During a typical 'long' session of one hour and forty-seven minutes, involving the edition of a large catalogue, consultation among users by multi-terminal exchange of messages, and the creation of two new files, text editor usage was 17.72 seconds and DIRAC execution time was 40.26 seconds.

We learned three major lessons from this experiment.

First, we found that the additional load on the time-sharing system was quite acceptable even in simulated situations where the astronomical data-base was interrogated and updated simultaneously by two or three terminals in addition to the main station at Dearborn Observatory. Even when the system was running at its full capacity of 60 terminals on-line, use of DIRAC on a number of these did not significantly degrade performance. This seems to show that a network such as the one hypothesized on figure 3 is quite within the state of the art in terms of operating system support, even when the environment includes batch and other time-sharing activities besides data retrieval.

Second, we could not observe a 'training period' of any significant duration: the astronomers usually devoted their first session to learning the basic text-editor and DIRAC commands, and started doing meaningful research work on the next session. This was quite an unexpected result, and it caused the number of users to expand quite rapidly beyond our initially very cautious estimates.

Finally, user acceptance was also better than we had anticipated once initial skepticism was overcome. The terminal proved to be a valuable teaching aid, and an astronomy class started to compile a galactic bibliography in DIRAC format; we had not anticipated this type of application. In another typical instance, an astronomer (Dr.Wray) interrogated the galaxy catalogue to extract a sub-list of all irregular galaxies for which radial velocities were not available, and used it that same day to prepare an observing schedule for the observatory's 40inch telescope. This particular use of DIRAC exemplifies best the advantage that can be drawn from new computing techniques in a research environment; this particular question, naturally, COULD have been solved by hand; but in itself, the simple fact that this study has never been undertaken in spite of the fact that the catalogue in question had been on observatory shelves around the world for years, seems to indicate

that the psychological investment required to initiate interrogations of this type is much lower under the network concept than it is under the constraints of a retrieval center operating by mail and, a fortiori, under the constraints of the isolated astronomer working with pencil and paper.

VI. COST ANALYSIS

This psychological factor is frequently overlooked when the problem of information retrieval is discussed among scientists. The question of cost is always posed - and rightly so ! - in terms of the obvious cost of hardware, and the not-so-obvious cost of software; but it is rarely balanced against the CURRENT costs of missing information, the cost of duplication of effort, the cost of slow and inadequate communication among researchers, the cost of the time lost in menial tasks by experts who should devote all their attention to data validation and high-level analyses. Very frequently for instance, astronomers will say: 'The cost of maintaining the data about a particular star in a direct-access data-base cannot be justified when we do not know if this particular record will be accessed more than once a month'. The answer to this, of course, is that no astronomer really knows how frequently this particular record may be needed by himself or his colleagues. No astronomer is even in a position to enumerate all the contexts in which the data about this star might be relevant, and no astronomer can answer the question: 'How much does it cost NOT TO HAVE the relevant information available when it is needed ?'

Assuming now that we work with users who have overcome this mental block and can grasp the usefulness of the machinery in the non-mathematical area, then the matter of cost-effectiveness can be settled in fairly simple terms. We have mentioned earlier the type of inquiry the ASP proposed for the 'data center' they were considering. Presumably these inquiries would be sent through the mail, and the center would operate in a manner similar to MEDLARS. In contrast to this, we tried to evaluate the cost of extracting and printing a number of copies of a significant subset of the Bright Star Catalogue from a remote terminal. The problem is to print offline, on a high-speed printer, a catalogue of all F-type stars that are the primaries of a Southern visual double. The DIRAC command the astronomer types is:

```
Delta < 0 AND mul2 ¬=0 AND Speluml >= 300 AND Speluml < 400
```

This command reflects a standard astronomical coding scheme that we have described elsewhere (6). The retrieval system found 192 such stars. Another command will now create an image of the catalogue under WYLBUR, the interactive text-editor with which DIRAC interfaces. This list of stars, with their parameters, is then available for punching, editing, or printing off-line. It could also be typed at the terminal or displayed on a scope. We generated 25 copies of this special catalogue on a high-speed printer with a final unit cost of fifty cents. The total time spent at the terminal by the astronomer had been six minutes, and he had spend 1.4 minute of computer time.

The trick in achieving this level of cost-effectiveness, of course, is that no programmer was involved in the interrogation. While the 'data center' concept would presumably include a large programming staff, and would maintain its own computer, we wanted to point out that a network could utilize an existing time-sharing service and that reliance on non-procedural programming, as Olle (9) and others have often pointed out, opened the way to a new and much more economical type of user/system interaction.

Our main observation, then, throughout this experiment and in discussions with professional astronomers that will be reported elsewhere (7), is that time-sharing and generalized programming combine to provide a suitable environment for efficient and cost-effective scientific documentation systems. But at the same time we found that the potential users of these systems were still largely unaware of their feasibility. We cannot escape the conclusion that the development of information systems as research tools might be delayed (with a correspondingly high expense due to duplication of effort and use of obsolescent software techniques in the mean time) for psychological reasons. These reasons will only be removed when scientists become aware of the availability of very large-scale, very cost-effective and very reliable business-oriented networks later in this decade.

REFERENCES

1. Simon Newcomb, Reminiscences of an Astronomer.

2. Survey of Generalized Data-Base Management Systems, CODASYL Systems Committee Report, 1969.

3. Publications of the Astronomical Society of the Pacific (P.A.S.P.) no.81, 922, December 1969.

4. 'Astronomical data in machine-readable form', U.S.Naval Observatory Circular no.114, 16 January 1967.

5. Vallee,J.F., and Ludwig,H.R. 'The DIRAC Language: Concepts and Facilities'. Research Report no.1, Information Systems Group, Stanford University Computation Center, June 1970.

6. Vallee,J.F., and Hynek, J.A. 'An Automatic Question-answering system for stellar astronomy' P.A.S.P. no.78, 463, August 1966.

7. Vallee,J.F. 'Scientific Information Networks: A Case Study' Research Report no.2, Information Systems Group, Stanford University Computation Center, forthcoming.

8. McGee,W.C. 'Generalized File Processing' in Annual Review in Automatic Programming, Vol.5 Pergamon Press, 1968.

9. Olle,T.W.,'Non-Procedural Languages- The next step in Automatic Programming' in Annual Review in Automatic Programming, 6, 1969.

10. 'The large Data-Base: Its organization and user interface', transcript of a panel discussion at the 1968 ACM Meeting, Las Vegas. Data-Base no.1, May 1969.

11. Hynek,J.A., and Wackerling,L.R. 'Feasibility Study of an Astronomical Data Network', International Astronomical Union, Brighton (England) Triennal Meeting, Aug.20,1970.

DATA CENTER OR INFORMATION NETWORK ? A SYMPOSIUM

(Transcript of a meeting of the American Section of the International Astronomical Union Committee on Data Processing, Boulder, Colorado, 13 June 1970.)

Participants :

-Abt	:	Dr.Helmut A.Abt, Kitt Peak National Observatory (Chairman)
-Bixby	:	Miss Joan E.Bixby, U.S.Naval Observatory
-Duncombe	:	Dr.Raynor L.Duncombe, U.S.Naval Observatory
-Hynek	:	Dr.J.Allen Hynek, Director, Dearborn Observatory
-Kieffer	:	Dr.L.J.Kieffer, Joint Institute for Laboratory Astrophysics
Moore	:	Dr.Elliot Moore, New Mexico Institute of Mining & Technology
Neff	:	Dr.John S.Neff, Dept.of Physics and Astronomy, Iowa State
-Roman	:	Dr.Nancy G.Roman, N.A.S.A. Headquarters
-Sitterly	:	Dr. Charlotte Moore-Sitterly, National Bureau of Standards
-Strand	:	Dr.Kaj Aa.Strand, U.S.Naval Observatory
-Wackerling	:	Mr.Lloyd R.Wackerling, Astronomy Department, Northwestern U.

Names preceded by a minus signs are those of the participants who have returned an edited version of their comments.

22

DATA CENTER OR INFORMATION NETWORK ? A Symposium

(After presentations by Dr.Abt and Dr.Sitterly dealing with existing catalogues and with the feasibility of an astronomical data center, Mr.Wackerling describes the time-sharing experiment)

- Abt : Well, maybe it's a good time to stop and let Lloyd take over, because he has tried experiments. Let me say one more thing, and that is, this data center could be operated by remote consoles, and this is what Lloyd is bringing up; because it's possible now to have remote consoles for a central computer, have these consoles in various observatories and have essentially instant access. An experiment along this line is what Lloyd has been talking about.

- Wackerling : I would like to start by shifting concepts. We talk about experiments pointing to the development of a North-American astronomical data network, INFORMATION network if you will. The Northwestern University Department of Astronomy and the Stanford Computation Center have just carried out a two-month joint experiment for the purpose of trying to evaluate in a preliminary sort of way how one goes about the creation and use of a large data-base by such a proposed network.
We used the 360/67 of the Stanford University Computation Center's Campus Facility and we put on various catalogues. Created them. Updated them. And queried them. I have samples of the work which we did, which I did just on Monday. We put on the Reference Catalogue of Bright Galaxies, this is the one we worked most with. We put on the Dearborn Observatory version of the Bright Star Catalogue, we put on the Warsaw Preliminary Supernovae Catalogue, and also created various sample bibliographic files and worked a bit with them.

I would like to tell you the scenario we have in mind for the creation of such a network. One might think of the establishment, over a period like two years or so, of a sizeable data-base, and one might think of the creation of the data-base being carried out by those people who are rather small in number, who are right now, currently maintaining and updating sizeable files of astronomical data, the double star files, all the files of the Naval Observatory, de Vaucouleurs' galaxy file... one can think of several examples. And, at the end of an initial period when a data-base of significant proportions has been established, one thinks of making the network hardware and fully-developed software available to use by people at large, and essentially the entire North-American astronomical community.

I would like to give you just a few remarks on the concept of the network. We want to stress the NETWORK concept, with the eventual active participation of the entire accessible population of astronomers, as opposed to the concept of the single data center, the single, large data CENTER, which is taking the responsibility for: acquiring, organizing, validating, verifying, and updating the wide variety of astronomical data files, and also on top of that, the responsibility for supporting the hardware and developing the sophisticated software technology which will be necessary to make this service available to the remote users. And in this connection it might be instructive to have the testimony of the people who are already in the game as to the feasibility of such a center, data center. There are data centers in existence, not in astronomy - thinking of other fields. And some people, more knowledgeable than I, think that there has not been created a data center which has worked.

(Some existing data center experiments, like MEDLARS and SPIRES/BALLOTS, are critically discussed).

- Wackerling : Another point is that the concept of regional computing networks is one which is - making use of a large computing facility - is one which is beginning to be acted on right now in the U.S. You have the University of California going on a state-wide net, well, it has plans to go on a state-wide net, and I think there are plans in other states to go on state-wide computing nets.

- Neff : Iowa has had one running for two years and it is very successful.

- Duncombe : North Carolina has had one for at least a year.

- Wackerling : And I think perhaps the next ten years will see great changes on network lines, and perhaps great changes in individual computation centers as we now know them, because there isn't enough talent to go around in maintaining the computation centers, but anyway...

And finally I wonder if you would like to hear described the software used, I'll try to do that quickly. The concept, then, that we have in mind, is the establishment of a large data-base at a central computer installation and its use in an interactive mode in a time-sharing environment, from remote locations, by anyone interested.

- Roman : Can you be a little more specific as to what you have in mind ? I understand all the words, but I don't understand...

- Wackerling : May I tell you what we used ?

- Abt : Can you give an example ?

- Hynek : No, what she means is: It's interactive: what do you mean, interactive ?

- Wackerling : I mean you sit down at a console and you talk to a computer.

- Hynek : And also point out the fact that the expert in charge at the Naval Observatory for instance, or whatever their responsibility is, could update the data-bank, but only that responsible center could.

- Wackerling : I wanted to come to that.

- Hynek : Well, fine.

- Roman : You say you interact with it. I don't understand in what way you interact. You ask it for data, and you get data, and then presumably you consider this is not really what I want, is that what you mean by interacting ? Or you ask it for data and you don't like it, and you tell it: the data you have in there is wrong, you should replace it by thus and so.

- Wackerling : Yes.

- Hynek : That's right.

- Roman : I would consider that somewhat dangerous.

- Wackerling : You've got to have update access to it.

- Abt : All you're doing is entering in a list to the expert, and the next time he looks at that particular group of things that has to be done, you have given him a comment that says: 'Look at the data for HD 1554, it's wrong. ' And he'll look, and re-evaluate it.

- Wackerling : But one has the picture of the double-star expert ...

- Strand : There we have the problem from the very beginning. People will think that when they find a mistake they can immediately correct the data in the bank. In regard to the Double Star Catalogs, we have insisted on people mailing us their corrections and these would not be incorporated until we had an opportunity to verify them.

- Hynek : Exactly. Without that you're opening Pandora's box. It's just a chance of all kinds of errors coming in.

- Neff : The security of your system, that puts some constraints on your hardware.

- Wackerling : The system in existence provides three levels of access to the data-base, which I'll come to shortly. So anyway, one envisions at every remote location a terminal and a scope for extraction of data. The machine we run on is a 360/67. It uses OS/360-HASP as an operating system. We use the Stanford time-sharing submonitor, which is named ORVYL, and the submonitor is relied on by the language which we use, which is called DIRAC. DIRAC was developed by Dr.Jacques Vallee, who was at Northwestern, he is currently manager of information systems at Stanford. DIRAC interfaces with and is driven by Stanford's interactive text-editor, which is named WYLBUR, and you also have available a terminal-to-terminal communication under a system named MILTEN, that is, any one signed on to the network can talk to anyone else signed on to the network. You also have the use of a language named ALTAIR, which has been described by Hynek and Vallee in P.A.S.P. some years ago, which answers questions put to it in a kind of English, I know you would like to say 'English'... it's approximately English. This answers questions on properties of stars, bright stars at the moment.

- ? : Can you give an example of conversation ?

- Abt : You can just quote it.

- Hynek : If you want just an example of question, you could certainly ask: 'Of all stars brighter than a certain magnitude, and earlier than G5, how many are members of visual binary systems whose space velocities are greater than fifty kilometers per second ?' You could do that.

- ? : Would you ask the question in exactly that form ?

- Hynek : Just about.

- Wackerling : And ALTAIR would come back and tell you if it didn't recognize any of the words.

- Hynek : I forget now whether you would use the word 'double' or 'binary', but obviously...

- Abt : It would be possible to have a language such that the computer would recognize the word 'double' or 'binary'.

- Hynek : That's a later edition, but in ALTAIR it didn't. If you said

'binary' and it only knew the word 'double', then it would come back and say : WORD UNKNOWN.

- Abt : OK, so you talk to it in English.

- Hynek : You talk to it in English.

- Abt : And it replies in English.

- Hynek : The whole point, however, in an advanced system, is essentially not to have the astronomer rely on a programmer to phrase the question. He gets down there and types out the thing, and the system essentially CONSTRUCTS THE PROGRAM to answer that question.

- Abt : Because it's a type of question you ask that is fundamentally so simple; because you're stating a magnitude or a distance, and you're asking is it greater than, or less than...

- Hynek : Yes, the ALTAIR system gave you numbers and ratios and percentages, things like that. If you wanted to prepare an observing list of things, you just made out the... It spoils it, in a sense !

- Wackerling : In the example I showed you when I loaded DIRAC, I said I wanted to query the galaxy file, and then I asked it - this was, incidentally, Vaucouleurs'reference catalogue of Bright Galaxies - I asked it for all the galaxies for which a Yerkes type existed, in one of Morgan's two lists, it extracted these; I said, 'retain that sample, I want to ask more questions', I then asked for all the A or F-type galaxies of irregular form, and it picked up those, and I then said, for how many does a radial velocity not exist, and it picked up those, then I said, 'print me out the Yerkes types', and whatever data I wanted to extract, I forget what they were, but this is the kind of thing that one can do, I have more examples with me, which I'll show you.

- ? : How long did all this take ?

- Wackerling : Oh, that particular one, that took part of an afternoon.

- Hynek : Well, for the whole thing, but the actual scanning ?

- Wackerling : Oh, the galaxy catalogue is a rather sizable file. The time it takes to scan it varies depending on how the current use is on the machine, but typically, during periods of light use, perhaps three minutes to scan the file. During periods of heavy use in the Pacific afternoon, it took, oh, perhaps twice as much.

(Machine size and storage capacity are discussed, as well as the cost of peripheral hardware).

- Neff : Let me ask you, if I wanted to ask about a star that is known as Alpha Lyrae, how would I ask it ?

- Wackerling : There is one thing that I didn't say and that Helmut brought up. File number ONE should be a file of names of astronomical objects. That's something we can start to work on today, you go into the file, you say the name field contains 'Alpha Lyrae', and, bang ! You get the names of Alpha Lyrae, and number 2 gives you the name that you would use in querying the system.

- ? : But it (increases the size ?) of your little system.

- Wackerling : I created a sample one from the Messier Catalogue.

- Abt : This is what I mean. For instance, in looking at the old literature, you find a lot of Lalande numbers, which are no longer used. One of the first things you have to do is cross-reference all these catalogues.

- Strand : And how about Wolf numbers ?

- Abt : They should be cross-referenced too.

- Neff : Have you been trying to look up Wolf number 59 ? (laughter)

- Abt : If you wanted identification, you might automatically list the common designations for stars, such as the HR number, the BD or CD number, CPD number, the HD number, the name and perhaps the GC number. A more extensive list, such as Wilson and Wolf and Ross numbers, could be printed by request.

- Strand : That's a big file right there, the identifications alone ! Which, incidentally, I would like to have a copy of.

- Hynek : Maybe we should start...

- Roman : You need some of the old things.

- Abt : So that, anything you do along these lines, if you only stopped at intercomparing the catalogues, and printed out what Mount Wilson number corresponds to what BD number, you've already done something useful.

- Wackerling : Somebody mentioned the organization ...

- Strand : It's going to create a big center but not very many great astronomers, you know. The astronomers are going to rely on centers for everything.

- Wackerling : The astronomers will be relying on themselves to maintain the data-base.

- Hynek : The whole is not greater than some of its parts !

- Roman : It seems to me that all you've been talking about is how you'd have a nice, convenient access to the data center, versus sending a question and getting back an answer some time later. I don't see how much you're going to get out of the data center unless you have data put in first.

- Wackerling : It's the PEOPLE who make the data center. I've been talking the way I do because of things I've been hearing and the things I've been reading, especially the COSATI reprint.

- Roman : I don't understand how you avoid the data center in your system.

- Wackerling : The data center is a machine, it's not a staff of people.

- Hynek : That's the difference.

- Roman : Somebody has to put the information in.

- ? : Somewhere on this computer there has to be a set of disk packs, or if they are not on somebody has to put them on.

- Hynek : Let's suppose that Observatory Q is responsible for.. double stars, let's say. They're a staff of professional people located in Oshkosh or some place. They are the ones who are responsible, and only they, for updating this thing. Now, on their terminal they create the catalogue at the data center, but the center is not people, it is a machine. In other words, the double star experts don't have to be housed in this center. That would get terribly top-heavy.

- Roman : That's a trivial problem.

- Hynek : What ? To have all the double star experts, all the galaxy experts, all the...

- Roman : No, where you house your experts is a trivial problem. You have to have a central file of information.

- Hynek : That's the machine, the disk.

- Roman : But how do you compile your data ?

- Hynek : That's the responsibility of the people who have that. Suppose the IAU decides that astronomer Z and his group are responsible for radio-sources. Then only the people responsible, given that responsibility, have the right and access to the central data-bank. And they...

- Hackerling : In update.

- Hynek : In updating it, yes, in that category. They could not come in and mess up, say, your double star catalogue.

- ? : What happens if the observatory responsible for updating the double star catalogue decides they don't want to do it ?

- Abt : That has happened in the past, but you give it to another observatory, such as the transfer of the Jeffers double-star catalogue from the Lick Observatory to the U.S.Naval Observatory.

- Hynek (To Miss Roman): I think it's a matter of semantics, Nancy, really, what you mean by 'Center' here.

- Abt : I would like to open a big discussion. I detect there is a lot of questions here you'd like to ask.

- Strand : I think you should allow Dr.Duncombe to say a few words here, since we have had such a center longer than most.

- Hynek : That's true.

- Abt : All right, may I interrupt this and ask Ray to make any comments he'd like.

- Duncombe : Thank you, Helmut. We've had some experience in this, over the last few years, in the role of a data center in certain aspects of data. We have tried to assemble star catalogues in machine-readable form, these data being distributed in Europe by the Astronomical Rechen-Institute of Heidelberg, Germany, and by Her Majesty's Nautical Almanac Office, Sussex, England. In cooperation with COSPAR, we hope to establish a data

exchange center, probably located in Paris, for ephemerides for space research, star catalogues, and observational data generated by the various laboratories involved.

- Abt : You told me, I think, of two, I thought they were separate. One was a catalogue of catalogues available on tape or cards, and the other is a cooperation between IAU Commission 4 and COSPAR for a Committee of which you're Chairman, formed to consider a catalogue for space observation, which is mostly short-lived objects. Is that wrong ?

- Duncombe : I think something slipped in the interpretation.

- Abt : Then there is another thing here, in which there is a proposal for a center in Paris which would direct interested people to where data can be obtained.

- Duncombe : This is exactly the point I'm speaking to.

- Abt : But, this center would not necessarily store the data itself.

- Duncombe : This is precisely the point.

- Abt : Alright.

- Duncombe : This Center would act as an information center to direct users to the data groups that are generating the data, or that already have the data available. In other words, the group, the office in Paris, would NOT store all the data, it would merely act as a clearinghouse for requests for data, or requests for observations and refer them directly to the groups who have the expertise in either making the observations or analyzing them, and so forth.

- Abt : How does this differ from just an updated version of your Circular 114 ?

- Duncombe : In Circular 114 we refer only to the data that U.S.Naval Observatory has.

- Abt : Oh. But it could be expanded to say that these are the known catalogues available on cards or tape, and they can be obtained from - a given observatory.

- Duncombe : Exactly.

- Hynek : How about compatibility of such data ? Suppose I write to Paris and want a certain catalogue. Don't I have to have the proper computing machine so that the data, the format is compatible ? Or can that all be straightened out ?

- Duncombe : You get a format of the data right along with the data.

(Compatibility and exportability of data are discussed)

- Abt : I could see difficulties, in the sense that identification may be different, you may run into identification problems.

- Duncombe : The point I'm trying to make is this. We refer people to the particular group that has expertise in that area. Here you are assuming that all of those various groups are going to introduce their data into one large data-bank, for YOU, and MAINTAIN it for you, so that people can go to one place to get the data. But you're making more work

for the people who are the experts in various areas, in putting the burden on them to keep your data-bank up-to-date. It becomes their responsibility, not yours.

- Hynek : All right, don't they have that burden right now ?

- Wackerling : They have it already.

- Duncombe : They have this burden right now but to a limited extent. When someone asks me to send a copy of some particular catalogue, I I tell them, "All right, I have a version, and I know there are errors in it, but you're welcome to it under those circumstances". How are we going to warn the user of this data bank ? He goes to that and he assumes he is getting the most up-to-date, correct information available.

- Hynek : That would also have to be indicated.

- Duncombe : Well, all right, this is a variation on what you proposed.

- Hynek : Yes, of course I should point out, as Lloyd pointed out, that this thing works on one continent, the North American continent, I don't see how you'd have a terminal in Paris connected to a universal data bank some place.

- Duncombe : Oh no, this is perfectly true, but this is an objection that might be raised.

- Strand : The process of verifying the data will be slow. By the time this is done the user might no longer be interested.

- Roman : I just want a clarification: Is this a place where you could send a list of a hundred stars and, say, they give you the photometric observations of these stars, or would they only give you all the places that had photometric data ?

- Duncombe : I believe that initially, all they will do is tell you where the photometric data is.

- Roman : Which is not really very much help.

- Duncombe : That's the problem.

- Roman (laughing) : As you know !

- Duncombe : The second objective, I believe, of this assemblage of data, in one place, is that the user can query it and get all of the photometric data on a particular object.

- Abt : Now, what you've been talking about and what Lloyd has been talking about are complementary, because Lloyd has been talking mostly about this, and you've been...

- Duncombe : Yes, but let me proceed: I see no difference as far as objective one is concerned, between what you're proposing and WHAT WE DO NOW among centers of expertise...

- Hynek : TIME is perhaps the difference.

- Duncombe : Except that you're putting a greater burden on the people who supply the corrections of all the material, in order to keep your data-bank up-to-date. Your second objective to have all of the data

available for interpretation is excellent, but how much is it going to cost you ? How much does it cost you, for instance, to keep one item of information on a star, in storage, and have it accessed, say once every two years ? How much is it going to cost you to retrieve that ONE BIT of information, and how much does it cost you to analyze data ? I have had HORRENDOUS experience running through files of star catalogues to try to find different designations for the same star, among star catalogues. It takes a TERRIFIC amount of time to do this type of analysis.

- Abt : Once you've done it for a star, why should somebody else have to do it for the same star again in three months? There is a question in the back...

- Kieffer : I don't want to interrupt, I'm somewhat of an outsider, but I have the distinct feeling that you're all going down a very strange path. You're very concerned about computers and computer use, and yet I don't think that you really know exactly what you want to do with them. Let me say this about that sort of problem. I operate a center which has gone down this road in a modest way, and there are three seriou problems. Up to this point, as far as I know, large data-base systems have al failed. The reason that they have failed (these are not my particular reasons) are first of all, goals are not well-defined and they change. Second, changes in hardware, and finally changes in software. Now, all three of these things seem to combine, to absolutely SINK large data-base systems. These problems should be carefully analyzed for the system you're proposing. If you really understand what the basic information is that you have stored in the data-base, and exactly what people want to know about it, it's absolutely MUCH BETTER to print it out and send it to people rather than have every one continually query your computer. It's EXTREMELY expensive. My understanding is that you are operating on very modest budgets. You ought to recognize that from the beginning, and not spin your wheels concerning very sophisticated hardware and software. I have a very good idea of what software costs, because we have to generate a lot of our own, for a much simpler system.

- Abt : Yes, that is a good point. In defense, I can say the astronomers have been putting together catalogues for hundreds of years, and they have a lot of experience, and it's not necessarily expensive. I could guess how how much it costs...

(Discussion of the amount of work involved in catalogue preparation. Dr.Abt points out it has not been a very large expense in the past)

- Kieffer : I didn't say that. What I'm saying is that if you really want to use a sophisticated computer system, you really are making an order-of-magnitude jump from the kind of thing you're talking about, which is, I think, still the best way to do things, in many ways...

- Abt : You're saying, this is expensive.

- Kieffer : No, I'm saying that instead of producing a hard-bound copy ONCE for people, you have this data all stored, either on disk, tape or whatever, and continually search that to get little bits of information, that's something that is orders of magnitude more expensive, and that's what I gather we have been talking about.

(The cost of the Stanford/Northwestern experiment is discussed, but the cost analysis was not yet available. It is presented below in the section 'Analysis of an experiment with direct-access

astronomical catalogues')

- Hynek : Well, the gentleman in the back of the room said that even before a hard cover is out, it's out of date, and you can't update. Do you know of any catalogue that is perfect ? And doesn't need updating as soon as it is out ?

- Roman : Yes, but you're going to have that trouble anyway.

- ? : Continually. You're continually updating catalogues.

- Hynek : It's a lot easier to update the data bank than it is to print a new catalogue.

- Abt : I could see...

- Strand : We're going away from the subject, of how much it is going to cost, this is what I'd like to know RIGHT NOW.

- Hynek : Well, we don't know the answer to that as yet.

- Strand : Well, then don't propose something unless you want to have it, unless you have some kind of an answer of what the cost will be. On the basis of this, let astronomy in the United States decide whether NSF should spend, for instance, two million dollars on such an astronomical data network, or whether the money should be spent on research. This is what it comes down to.

- Hynek : You're absolutely right, this is what it does come down to.

- Duncombe : This is what it really comes down to.

- Hynek : And I think that is where the philosophy of this thing has to come in. NSF is supporting a fantastic amount of material for data GATHERING. Now, shouldn't there also be an AURA for the data processing and data use ? I mean, what's the point of putting millions of dollars into getting more and more and more data, if it becomes increasingly difficult to use it ? This is what we are faced with.

- ? : You would have to demonstrate that that's the case. You have to demonstrate that a typical astronomer is missing data he should use in solving whatever the problem is...

- Abt : Well, I can think of a very good example of a problem which suddenly became urgent about fifteen years ago and the data center would have answered instantly or in a matter of minutes, and that is: what cepheids are known to be in open clusters ? Because the realization that there are cepheids in open clusters had quite an impact on astronomy about fifteen years ago, and that could very easily have been answered.

- Roman : That probably is still not beyond the reasonable capability of answering by hand with an existing catalogue.

- Abt : Well, it was done by hand, but it took months to do.

- Roman : A more serious question is to take a list of eighth magnitude stars and look for their spectral types or their photometry...

- Hynek : Or their space motions.

- Abt : Can anybody give an example of really urgent problems that may have a big impact on astronomy but cannot be done by hand, probably like asking the number of cepheids in clusters. Something you really have to do with a computer ?

(The discussion turns to other retrieval systems)

- Abt : Well, the astronomers of the world now turn out something like 1500 and maybe 2000 papers per year. It's increasing over time. How many of these things, how many of these papers can any one person read, or scan, or even remember, and even in a relatively narrow part of astronomy ?

- ? : That's what the AJB (Astronomisches Jahresbericht. Ed.) Center...

- Abt : Yes, but the AJB is very poorly set up to recover all the data, because you have to look in every single volume. If you want to find out what's known about HD-so-and-so, do you want to sit down and look through seventy volumes of the Jahresbericht ?

- Bixby : When you consider the cost of maintaining a computer terminal which you're perhaps using once a month to ask about HD-something-or-other, as compared to the cost of sitting down with the Jahresbericht, there is just no comparison !

- Abt : But the point is, people don't do it, and therefore the...

- Bixby : But they do it !

- Abt : They do it ? Well, yes, in part, but it's more likely that they overlook past things. And I think our research will gradually deteriorate because people will not be aware, for instance, of the fact that when they work on a star, it may be a double star system and therefore may have a composite spectrum.

- Roman : The main thing is, you can't find out what's known on HD-something-or-other unless HD-something-or-other was sufficiently important to have a paper all to itself. If it was just one of a hundred stars in the paper you're not going to find it in the Jahresbericht.

- Hynek : Yes, that is a good point.

- Abt : Ray, you...?

- Duncombe : I was about to say: THERE IS NO SUBSTITUTE FOR DILIGENCE AND INDUSTRY.

- Hynek : Hear, hear !

- Duncombe : Even your proposed astronomical data bank won't do it.

- Sitterly : The EXPERT cannot be moved out. I think this is very important in all your plans. His careful planning is required for the success of any data center. Reliability of data given out to an inquirer is the important factor.

- Hynek : Oh no, you're absolutely right, you don't... But the point is: is astronomy going to be left in the rear, and the medical profession is doing this, the legal profession is...

- Sitterly : But they aren't successful.

- Hynek : Maybe at the present, all right; now does that mean that ten years from now they won't be doing it successfully ? Maybe we should profit from their mistakes !

- Abt : Do you think it's too early to do this ?

- Sitterly : I do. More groundwork is needed.

- Abt : You think we should wait ten years or so ?

- Neff : I think a minimum thing that could be started, in which I take there was a consensus, would be a star index: just a simple-minded star index.

- Abt : Like Lloyd in Illinois starts...

- Neff : You know, where you have all the names...

- Hynek : All the names.

- Neff : All the names of the stars known, and what is the information that's available.

- Roman : And right ascension and declination.

- Hynek : yes.

- Abt : How far would you go ?

- Neff : Well, any star that has a name ought to be in it.

- Roman : Name or number.

- Strand : We have names of stars as faint as twenty-first magnitude.

- ? : If anybody has bothered to name it, it should be in the index.

- Neff : Maybe I'm not posing the question correctly, but it seems to me there was a consensus that a star index of some form would be valuable.

- Hynek : Yes, but it's one of these things where there's a big ditch, you can't jump a quarter of the way. You have to make the total jump, and you either have a total data system that will be used for all sorts of things, or you don't have a data system.

- Bixby : Can't you start small, by having, say, a continuation of what's already done with star catalogs in a more organized form: One group is working with star catalogues, and another group is working with spectral types...?

- Hynek : That's exactly what we've proposed. AND WE WANT TO CONNECT THEM BY A NETWORK.

- Bixby : Yes, but this...

- Roman : It would encompass the whole world ?

- Hynek : No, no, no, well...

- Strand : You can do us a favor. We need identifications of stars from

various catalogues, and an astrographic catalogue, would you like to do that for us ?

- Hynek : Why, sure !

- Strand : That would save us a lot of money.

- Abt : That's a big problem, because different catalogues have positions that are not the same equinox. You can't just take a coincidence, you have to give a tolerance to the coincidence, and even then it's not reliable... (laughter)

- Strand : We know all about that. We have discussed it at the Maryland Astrometric Conference.

- Neff : I think that I would like to propose that we start with something we can do, preferably something small, as a pilot project, then, keep a careful accounting of what it costs...

- Abt : Yes but, John, can you be a little more specific, because we just had a pilot project.

(Confusion)

- Duncombe : Look, I'm not trying to sledgehammer this, all I will say is, you can make a computer do anything, it's how much it costs you to do it.

- Abt : And also, how much do you gain out of it ?

- Duncombe : Just get the balance: The benefit to you, against how much it's costing, and how many other demands there are for the same dollar. How much research..

- Roman : And the same expertise.

- Abt : Now, that will come in the proposal, because presumably there will be a proposal from Hynek and Wackerling and Jacques Vallee or somebody else, and they will estimate costs. Then the NSF or NASA or whoever is the recipient of this proposal will have to decide: Do we want to spend a certain fraction of money to do this, or fund some other projects. But, the only thing that we should discuss is the capabilities and astronomical possibilities of an astronomical data center, how important they are. So, if you have any suggestions, some additional ones, I'd be glad to hear them, because this Committee is going to terminate, and it will depend on the initiative of somebody to do the next thing. Any other comments ? Well, it's time to stop. Thank's very much for your time.

(Many people leave the room. The discussion continues among a small group)

- Abt : We've learned two things here: You've got to get some cost estimates, and the other is, you have to solve this problem yet.

- Wackerling : I disagree, we certainly do have in mind that problem exactly. That's one thing I wanted to remark, there is a catalogue...

- Abt : Like the planetary nebulae ?

- Wackerling : That's the format, that's very much the format that one thinks of. I did indeed type a sample of the planetary nebulae catalogue, to see what it looked like.

(Discussion of budgetary levels in Astronomy)

- Hynek : Well, the point is, it's like the old Pillsbury Flour advertisement: "Eventually... why not now ?" Eventually, some much more efficient way has got to be found to handle the astronomical data. Now, this may not be the way to do it, but anyway we've started thinking about it, that's the important point. Somebody will come up with a really...

- Wackerling : Trouble is, there are needs for this type of thing right now.

- Hynek : See, if we have a console at the Observatory, and I want to get a certain star, I can get the finder field with precessed positions. You wouldn't want to have a computer just for that, but...

- Bixby : It would be nice... I wouldn't see that you would use it just for access...

- Hynek : You probably wouldn't be using it as much as some of the people at the University of Wisconsin. It could be that some of the people who are generating the data in that particular field would not need it, because they got it there.

- Bixby : yes.

- ? : Now if you suddenly develop an interest in radio-sources, now you might want to have access to...

- Bixby : In some ways it might seem useful to copy the entire catalog and to work with your own computer. It might be more economically feasible and easier to adapt to a particular application than to try to work through a console.

- Hynek : On the other hand, that thing may get to the point where it was cheap to have it.

(Discussion of hardware available at Naval Observatory)

- Hynek : Well of course, she brings out an important thing, that they are specialized, so specialized in their data that they are the ones who are gathering it and USING it, they don't have any need to go and query a data bank for data they've already got.

- Wackerling : Right. But one envisions the Naval Observatory as being a 'participant' in the final...

- Bixby : The way things are run at the moment, we are generating data, and we are giving people copies. There have been times when it has been a burden on people to get a corrected copy of the data requested prepared for someone. This makes me wonder if a researcher would be willing to stop for three or four months to help set up a data bank. That's three or four months lost from the research.

- Wackerling : That's right, but people are doing this. Arne Slettebak does maintain a file of stellar rotational velocities, you maintain hundreds of records, and if this was presented to you as a tool, and it

were useful to you in your data compilation and your data updating, this is how one would hope to make it useful to YOU, you see ?

- Bixby : Yes.

- Wackerling : But, as I say, we have to make our case very much more clear.

- Bixby : Also, there is a certain element of trust that you have to have in the system. So far when presented with compilations of data most people generally try to do some checking or COMPLETELY check to see that that's a complete list and, well, many times you find it's not, and this kind of a system...

- Wackerling : But on the other hand, everyone does take Arne Slettebak's word that this is indeed the best stellar rotational velocity for blah blah blah, and if it's certified by A.Sletterbak November 10th, 1971, one perhaps trusts the data..

- Hynek : But this is something that covers the whole waterfront, I mean no matter...

- Wackerling : This is the point, you need CRITICAL data evaluation, a file of critically evaluated data.

- Hynek : Take any catalogue, take that supernovae catalogue that came from Warsaw, how do we know what its critical accuracy is ?
But if we were in closer communication with them, we could find out and test it, and check it. And heaven knows, even in the Bright Star Catalogue ... Well, we'll see what the problem will be... Can you drop me over at the Travelodge ?

ANALYSIS OF AN EXPERIMENT

WITH DIRECT-ACCESS ASTRONOMICAL CATALOGUES

J.Vallee

This report discusses the technical and economical aspects of an experiment with a prototype scientific network, which has been described above in terms of the contents of its data base and its user interface.

The experiment itself took place from April 2 to May 26, 1970. An IBM 2741 Communications Terminal owned by the Stanford Computation Center was shipped to Evanston and installed at the Dearborn Observatory for use with a dataphone coupler. Three major catalogues were created and stored on direct-access disk packs at Stanford. The Warsaw catalogue of supernovae was converted to machine-readable form in DIRAC labelled-input format for the purpose of the experiment. The Dearborn Observatory version of the Bright Star Catalogue was created directly from the existing tape through DIRAC's positional input processor, and so was the Reference Catalogue of Bright Galaxies obtained from the University of Texas. Emphasis was then placed on the creation of approximately ten more experimental files reflecting the personal interests of the users: These included a Messier catalogue, a galactic bibliography, etc.

Throughout the experiment, a careful account was kept of the following data:

1. The parameters of each session. 2. The computer time and service charges, broken down by category. 3. The telephone line charges. 4. The various other expenses such as travel and training.

The object of this effort was to identify qualitatively and quantitatively the parameters of the interaction between a group of scientists, non-programmers, located far away from a computing center, and a data-base with which they could only communicate through the generalized software itself.

It was of particular interest to observe the distribution of the sessions and the nature of the work accomplished; to analyze the distribution of holding times; to study the causes of the failures that occurred; to break down the cost of the experiment according to its main components; and finally to try and form a picture of the economics of a fully-implemented network. We propose to study these items in turn.

I. ANALYSIS OF THE MAN/SYSTEM INTERACTION.

A large time-sharing system, such as the one running at Stanford, develops a well-defined pattern of behavior as a function of time, number of users and mix of jobs in its various partitions. Before the development of DIRAC early in 1970, information retrieval applications had only minimal impact on this environment, and when it occurred this interaction was not measured. The release of DIRAC or of some similar tool to a potentially large user population would have repercussions on the system, and it was important to understand this phenomenon. This provided the main motivation for Stanford's support of a limited

data-base experiment.

To be meaningful, the experiment had to involve remote users with a significant time difference to the central computer. It was also of interest to select an application area that represented a new type of activity at Stanford. Thus, Dearborn Observatory, with its two-hour time difference to California, its expertise in astronomical catalogues, and its interest in computer applications in Astronomy, was an ideal candidate.

Most of the actual work of supervising the creation of the data-base and training newcomers in the use of WYLBUR and DIRAC was done by Mr. Lloyd Wackerling. Other astronomers - notably Mr. J. Mack and Dr. James Wray - devoted considerable time to the experiment. A total of eight astronomers took part in the two-month study, all of them from Northwestern. (Stanford University has no Astronomy Department).

One hundred and twenty sessions were logged. Users were asked to identify every session, to record the date, starting time (PST), total duration, CPU utilization, editing time; in addition, they recorded the mode of termination (Normal/Abnormal), the name of the user, the systems used, and their remarks. (Some of these parameters are printed out by the system during normal termination, but users do not usually save them as they rely on the monthly statement they receive from the accounting office). This information is recorded in Table I below.

We have presented elsewhere an analysis of the distribution of holding times and of the distribution of sessions during the day. We have also discussed the differences that are apparent between these graphs and the typical distributions observed on the Stanford system for other users. Figure 1 displays the distribution of the sessions of known duration for each day of the experiment.

The duration of the sessions and the amounts of computer time used reflect the nature of the work that was performed: the system was utilized essentially at three levels: for communication and consultation among users, for editing of the primary data sets (in WYLBUR) that were used in generating the DIRAC files both in creation and in updating; and mostly, of course, in querying the resulting files.

II. ANALYSIS OF SYSTEM FAILURES.

Because none of the users had experience in the operating system, and because access to programming languages other than DIRAC was not given to them, a situation was created that made failures especially critical.

TABLE I

wk	session no	date	logon hr:m	durat hr:m	CPU time (sec) ORVYL	CPU time (sec) Editing	t	Systems W	Systems D
1	1	2 IV	10:00	0:15	4.74	4.10	N		X
	2	2 IV	10:57	0:27	5.81	9.83	N	X	X
	3	2 IV	8:21	1:47	40.26	17.72	N	X	X
	4	2 IV	20:41	1:24	5.04	16.27	N	X	X
	5	2 IV	22:08	0:13	0.00	0.88	N	X	
	6	3 IV	10:01	0:21	0.00	0.05	N	X	
	7	3 IV	11:43	0:43	0.00	22.51	N	X	
	8	3 IV	12:57	0:43	0.02	4.72	N	X	
	9	3 IV	18:37	0:39	13.76	12.70	N	X	X
	10	3 IV	19:37	2:59	221.51	76.48	N	X	X
	11	4 IV	11:35	0:58	59.05	29.[illegible]5	N	X	X
	12	4 IV	14:00	2:11	329.56	48.92	N	X	X
	13	5 IV	14:41	?	?	?	A	X	X
	14	5 IV	15:38	0:24	8.28	4.52	N	X	X
2	15	6 IV	9:18	1:09	4.35	4.91	N	X	
	16	6 IV	11:47	?	?	?	A	X	X
	17	6 IV	13:57	1:09	?	2.37	A	X	X
	18	6 IV	15:42	0:27	0.51	4.28	N		X
	19	7 IV	12:10	0:36	65.76	4.39	N		X
	20	7 IV	14:40	0:15	0.06	0.14	N	X	
	21	7 IV	19:38	1:19	?	119.90	A	X	X
	22	8 IV	9:45	0:20	?	45.57	A	X	X
	23	8 IV	11:16	1:28	238.79	252.59	N	X	X
	24	? IV	?	?	?	?	N	X	X
	25	9 IV	14:23	?	?	?	A		X
	26	9 IV	14:40	?	?	?	A		X
	27	10 IV	9:00	1:43	208.88	33.70	N	X	X
	28	10 IV	11:27	2:04	?	?	A	X	X
3	29	13 IV	8:48	1:35	6.90	6.53	N	X	X
	30	13 IV	11:48	3:09	240.98	63.16	N	X	X
	31	14 IV	19:48	1:45	61.68	26.85	N	X	X
	32	15 IV	8:24	2:06	228.32	41.32	N	X	X
	33	15 IV	?	3:41	73.07	50.28	N	X	X
	34	16 IV	8:22	?	?	?	N	X	X
	35	17 IV	11:11	#1:20	?	?	A	X	
	36	17 IV	12:40	2:48	261.25	55.53	N	X	X
	37	17 IV	18:57	2:02	30.92	19.23	N	X	X
	38	18 IV	12:18	#0:25	?	?	A	X	
	39	18 IV	12:49	?	?	?	A	X	
	40	18 IV	15:07	0:01	0.00	0.00	N	X	
	41	19 IV	18:01	0:51	251.86	10.21	N	X	X
4	42	20 IV	8:21	1:67	40.26	17.71	N	X	X
	43	20 IV	11:24	1:53	8.08	7.57	N	X	X
	44	20 IV	19:57	1:39	0.16	4.02	N	X	
	45	21 IV	8:17	1:49	320.61	44.84	N	X	X
	46	21 IV	11:21	3:41	0.18	1.89	N	X	X
	47	22 IV	8:44	0:33	180.36	12.45	N		X
	48	22 IV	10:59	?	?	?	?		X
	49	22 IV	13:10	2:13	339.95	33.87	N		X
	50	22 IV	18:07	0:43	73.67	13.62	N		X
	51	22 IV	?	?	?	?	N		X
	52	23 IV	9:16	0:35	63.01	4.56	N		X
	53	23 IV	11:07	1:29	336.36	10.24	N		X
	54	23 IV	?	?	?	?	A		X
	55	23 IV	19:53	1:12	607.56	4.26	N		X
	56	24 IV	?	#1:00	#360.00	?	A		X
	57	24 IV	10:19	1:34	152.33	30.46	N		X
	58	24 IV	11:54	?	#40.00	?	A	X	X
	59	25 IV	11:13	#0:35	?	?	A	X	
	60	25 IV	11:48	?	?	?	N		
	61	25 IV	11:49	0:30	176.18	4.30	N		X
	62	26 IV	17:48	0:37	99.99	10.81	N		X
	63	26 IV	20:28	0:52	216.80	8.65	N		X

TABLE I

wk	session no	date	logon hr:m	durat hr:m	CPU time (sec) ORVYL	Editing	t	Systems W	D
5	64	27 IV	8:27	0:57	376.27	4.49	N		X
	65	27 IV	11:43	1:16	369.66	9.87	N	X	X
	66	28 IV	11:08	?	?	?	A	X	
	67	28 IV	12:58	#2:00	?	?	A	X	X
	68	28 IV	15:00	0:08	0.22	0.53	N	X	
	69	28 IV	20:04	1:34	256.32	22.31	N	X	X
	70	29 IV	8:18	3:44	139.34	22.04	N	X	X
	71	29 IV	12:05	3:11	237.63	13.83	N	X	X
	72	29 IV	20:12	2:14	481.26	13.91	N		X
	73	30 IV	8:25	1:20	20.35	43.09	N	X	X
	74	30 IV	13:32	1:53	318.66	20.90	N	X	X
	75	30 IV	19:37	2:69	22.66	13.86	N	X	X
	76	1 V	8:00	2:09	364.71	15.62	N	X	X
	77	1 V	11:52	1:07	78.47	28.60	N		X
	78	2 V	10:27	0:19	126.22	4.78	N		X
	79	3 V	10:25	2:36	880.80	13.35	N		X
6	80	4 V	8:10	2:01	357.93	32.70	N	X	X
	81	4 V	10:21	0:12	48.97	4.70	N		X
	82	4 V	12:26	#1:46	#120.00	?	A	X	X
	83	4 V	14:13	0:16	0.18	0.16	N		X
	84	4 V	14:33	0:12	0.19	0.16	N		X
	85	4 V	16:04	?	?	?	A		X
	86	5 V	9:37	0:09	10.92	9.75	N		X
	87	5 V	12:35	2:39	?	6.95	N	X	
	88	6 V	11:26	1:05	227.17	10.28	N		X
	89	7 V	8:41	0:28	84.04	4.42	N		X
	90	7 V	12:17	0:05	5.93	4.78	N		X
	91	7 V	12:32	3:04	84.62	113.46	N	X	X
	92	8 V	8:68	1:50	326.48	157.90	N	X	X
7	93	11 V	9:09	0:07	?	?	A	X	
	94	11 V	9:19	0:05	0.17	0.16	N	X	
	95	11 V	11:37	0:05	0.00	0.01	N	X	
	96	11 V	12:41	0:03	0.00	0.00	N	X	
	97	11 V	20:35	0:03	0.00	0.00	A	X	
	98	11 V	20:39	0:45	104.10	26.66	N		X
	99	12 V	12:38	1:36	69.72	18.75	N		X
	100	12 V	20:36	#0:20	#15.00	?	A		X
	101	12 V	21:37	1:59	485.24	25.78	N		X
	102	13 V	12:05	#2:10	?	?	A	X	
	103	13 V	14:16	1:13	33.26	43.07	N	X	X
	104	13 V	20:44	2:23	#180.00	23.18	N		X
	105	14 V	8:29	1:43	282.56	82.28	N	X	X
	106	14 V	11:43	0:27	0.16	52.25	N	X	
	107	14 V	13:03	1:34	?	41.32	N	X	X
	108	15 V	8:38	1:38	341.60	15.86	N		X
	109	15 V	11:32	#0:20	?	?	A	X	
	110	15 V	11:55	3:21	#7.15	50.23	N	X	X
8	111	18 V	8:24	1:30	324.99	14.39	N	X	X
	112	18 V	12:21	#2:05	#5.00	?	A	X	X
	113	18 V	14:33	1:01	156.60	4.94	N		X
	114	18 V	19:05	#0:05	?	?	A		X
	115	18 V	19:12	2:29	416.52	4.59	N	X	X
	116	19 V	8:28	1:37	243.38	4.79	N		X
	117	19 V	11:43	2:07	313.25	5.35	N		X
	118	20 V	14:43	0:29	0.18	0.38	N	X	
	119	21 V	11:34	0:48	26.27	7.65	N		X
	120	25 V	10:54	0:42	59.31	5.22	N		X

However, this afforded a good test of the ability of scientific users to solve technical problems pertaining to recovery of their data, and also led to useful suggestions to the system designers. In general, system failures did not cause major psychological problems, except during two or three days when a major flaw existed within the file system itself, and during one period in April when drum errors were frequent.

Out of 75 sessions in April, seventeen terminated abnormally. In May, there were nine abnormal terminations for 45 sessions. The major cause of failure appeared to the users as a loss of communication . On only one occasion did WYLBUR itself die and cause termination. On another occasion, a telephone operator's mistake was responsible for the failure. In the remaining cases the terminal 'hung' in ORVYL or while attempting to load DIRAC. A few sessions showed problems due to experimentation with ORVYL files at Stanford.

Of those sessions that eventually terminated normally, some had serious problems during which it was necessary to re-load the system. Under such circumstances the files and data-sets are not destroyed and the users are advised to save and either logoff or await recovery. Such problems are naturally unavoidable even in a stable system, and the experiment showed the users responding perfectly to the messages broadcast from the console. There was no loss of any sizable data-set due to system failure or to user mistake during the entire experiment, a very remarkable fact indeed. Much of the time between failures and recoveries was spent in consultation through the terminal-to-terminal communication system, which remained open in most cases.

Generally speaking, the major cause of failure was the combination of the large-scale testing of the DIRAC prototype and of the experimentation that was still taking place with the file system on which DIRAC relied. However, these failures made it possible to correct the situation much faster by providing a wider scope of field tests, and the file system reached a stable state in May.

Some of the failures resulted in loss of accounting information, reflected in the question marks in some columns of table I.

III. COST ANALYSIS OF THE EXPERIMENT.

The total cost of the experiment was approximately $10,000 - a figure that does not provide a true indication of the economics of running a full-scale information network, for various reasons that will be studied below. It does provide considerable insight into this situation, however.

Table II shows the actual expenses charged to the network experiment by the Accounting office. Although these costs were supported out of the development budget for DIRAC, a special account number had been obtained so that separate accounting could be shown. Of special interest in Table II is the contrast between the terminal access time and the actual amount of computer time used. Also of interest are the ratio of editing time to ORVYL time, and the rate of growth of the data base between April and May, from 2122 tracks/day to 4611 tracks/day. These figures , are probably a fairly good indication of the actual parameters in a full-scale network. We point them out with special emphasis because most proposals for the creation of such systems under-estimate these parameters and their rate of growth.

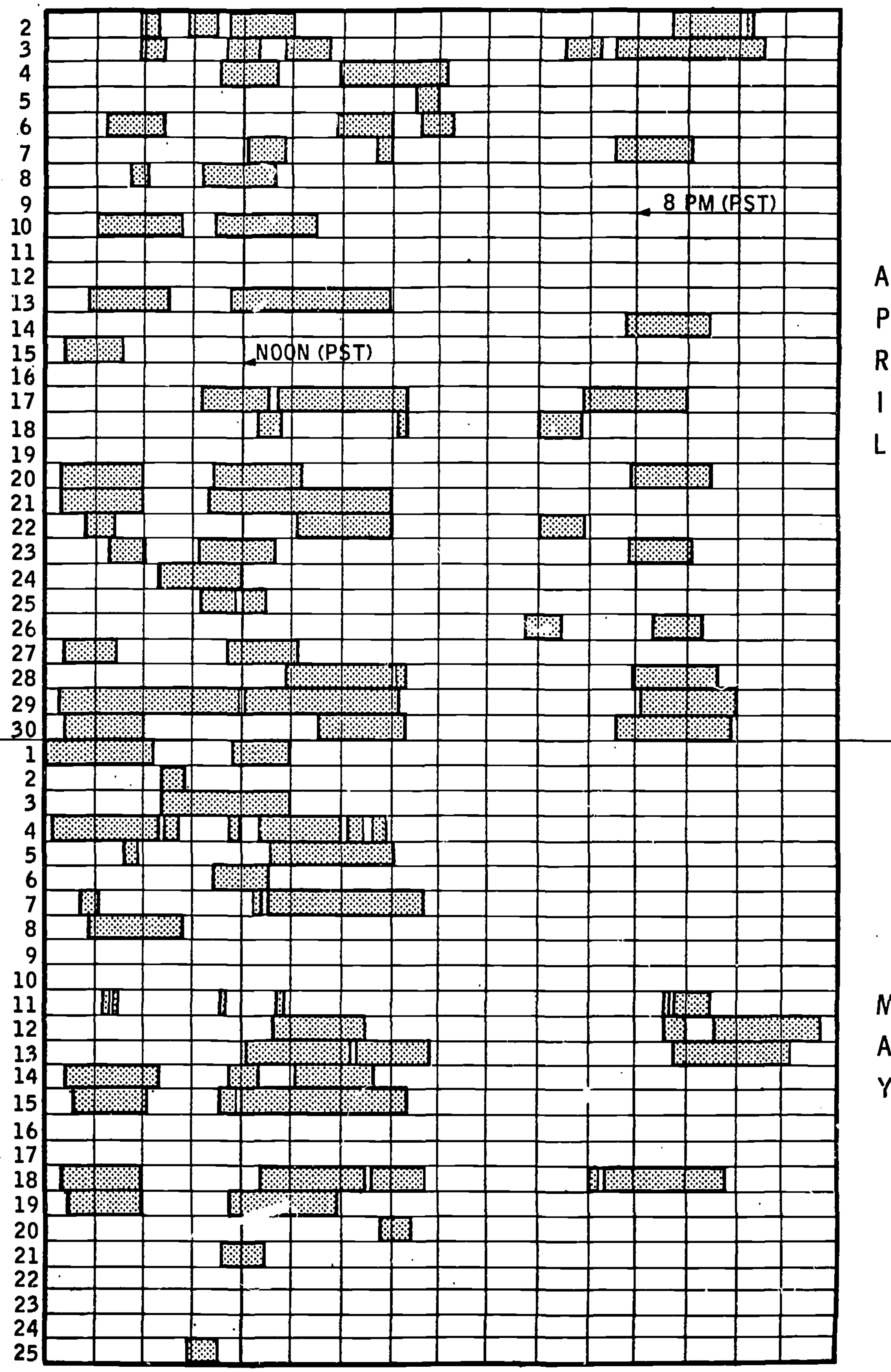

8 PM (PST)
NOON (PST)
A P R I L
M A Y
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

TABLE II: COMPUTER COSTS

	Terminal Access	CPU time		Disk Storage WYLBUR only	Total
		ORVYL	editing		
	(minutes)	(min.)	(min.)	(tracks/day)	$$$
April 1970	8364	208.3	64.3	2122	3,620.64
May 1970	4851	159.9	52.0	4611	2,956.36
TOTAL	13215	368.2	116.3		6,577.00

The evolution of these costs from the first to the second month reflects the fact that, while the size of the data-base grew rapidly, the actual use of the terminal was very much reduced, without as large a decrease in terms of ORVYL and WYLBUR time. This shows that during the second month, users made better use of the system at their disposal and that the ratio of CPU time to holding time evolved in a more economical direction. The experiment, however, was too short to establish the trend.

A fairly sizable travel budget had to be associated with the experiment. A preliminary meeting had taken place in December 1969 in Stanford. In February, Messrs. Wackerling and Mack spent two days at Stanford familiarizing themselves with the very first DIRAC prototype, and offered suggestions that were implemented before the experiment began. Between April 7th and April 11th, I visited Evanston to review the early results and demonstrate new features of the system. Later, Dr. Hynek and Mr. Wackerling traveled to other universities to obtain comments and suggestions from astronomers about the project. A final review meeting was held in Evanston on July 8 and 9. The total travel and training expense (adding the contributions of both Universities) thus amounts to approximately 18% of the total cost, and it is unlikely to be much lower in a full-scale network of this type, although the allocation to travel itself would probably be smaller compared to the training budget.

IV. ECONOMIC CONSIDERATIONS IN A FULL-SCALE NETWORK.

Within the scope of this study, we mean by 'full-scale network' an extension of the user population of this prototype experiment to include four or five other locations for which a pattern of behavior can be predicted. Presumably this population would include most of the catalogue experts in North-American astronomy, or would be organized so as to serve their needs according to the concept of distributed responsibility over the files that will be accumulated. This is in contrast to the predictions one might be tempted to make concerning a very large network offering subscriptions to universities and colleges as well as to research establishments, that could reach over one hundred users. The concept of such a network is certainly feasible, but it could be best considered as a second step, after the accumulation of a

meaningful data-base by a team of experts.

TABLE III

Wk	no.	comp.	duration hr:min	ORVYL sec	Edit sec	weekly averages: duration min	weekly averages: ORVYL sec	weekly averages: Edit sec	weekly averages: CPU sec	%CPU	%Edit
1	14	13	12:44	708.7	255.5	58.7	54.5	19.4	74.0	2.1	26
2	14	6	5:38	518.4	300.0	56.3	86.4	50.0	136.0	4.0	37
3	13	9	17:58	1154.9	273.1	119.6	128.3	30.3	159.0	2.2	19
4	22	15	21:07	2614.5	209.3	84.5	139.5	13.9	153.0	3.0	9
5	16	14	25:17	3672.6	227.2	108.4	262.3	16.2	278.0	4.2	6
6	13	10	9:22	1145.7	338.3	56.2	114.6	33.8	148.0	4.3	23
7	18	13	15:11	1504.3	338.2	70.0	115.7	26.0	142.0	3.3	[illegible]
8	10	8	10:43	1540.5	47.3	80.3	192.6	5.9	199.0	4.1	3

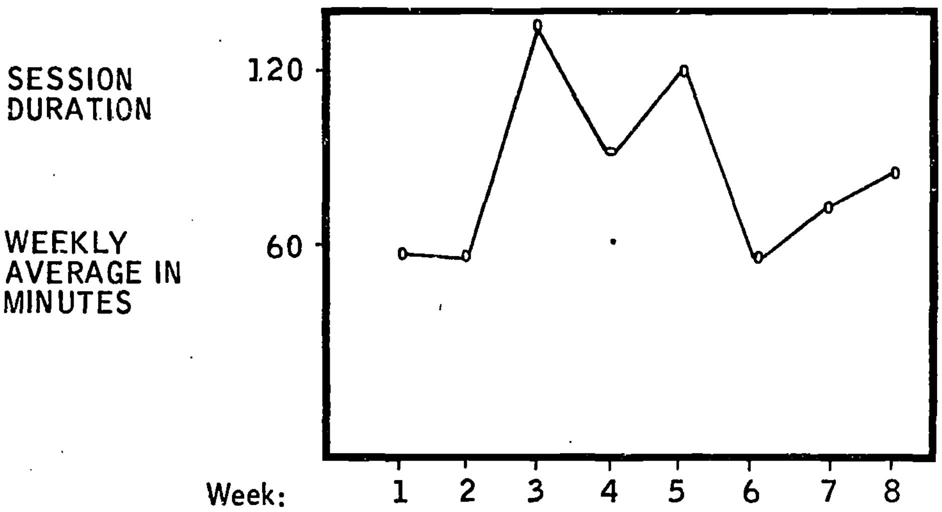
Table III above summarizes the statistical data for the eight weeks of the experiment, giving for each week the number of sessions, the number of sessions for which complete statistical data were available, the total terminal connect time, the total ORVYL and editing times, and the average time per session for these categories. Also given is the percentage of CPU time per session, and the percentage of editing time per session. It is of interest to examine these data on the following graphs. It becomes apparent that, while the average session duration fluctuated more or less randomly, the percentage of computer time tended to rise. At the same time, that fraction of CPU time devoted to text editing tended to decrease. If it is true that this behavior can be extrapolated to a long-term network performance, then certain steps could be taken to decrease the operating costs. We shall now review a few tentative ideas along these lines.

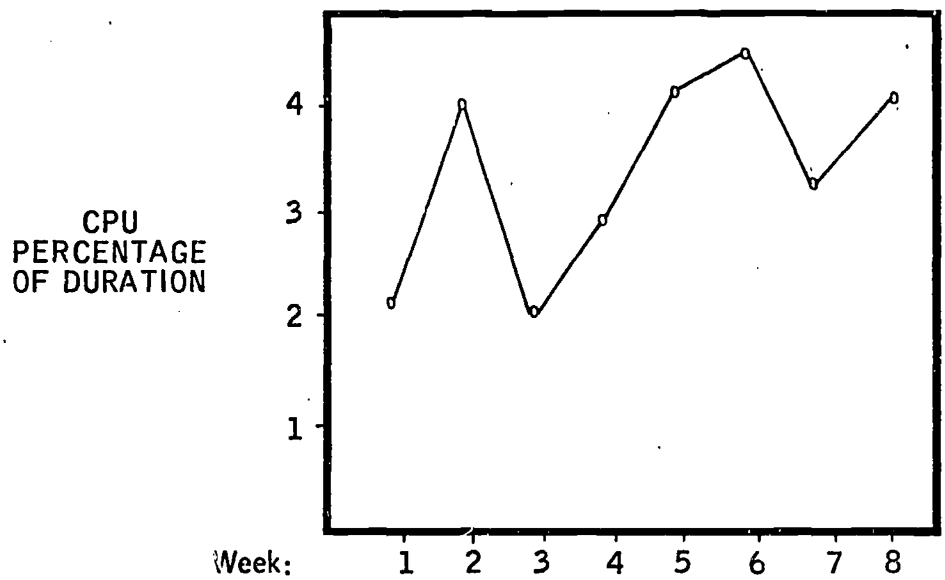
1. Location of central computer.

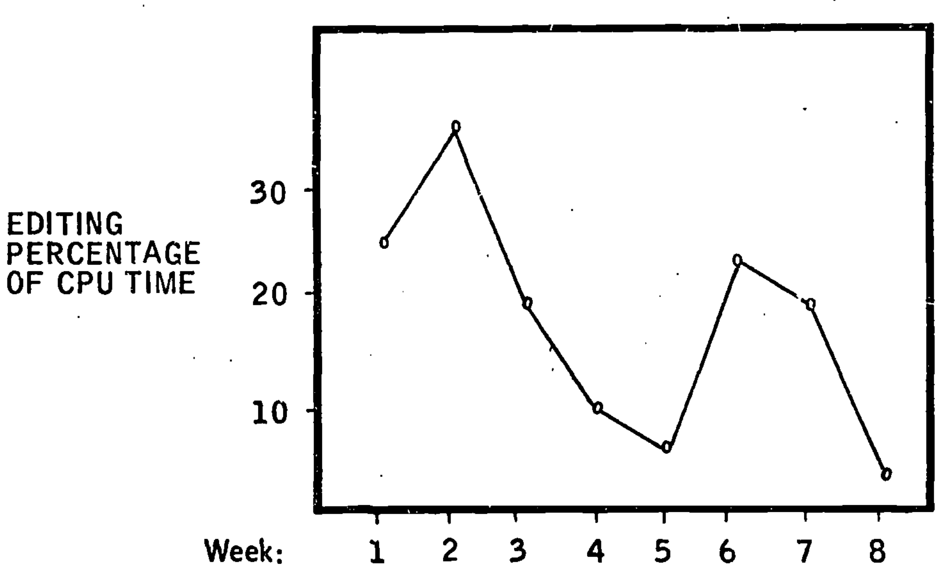
We assume that only one computer is available; given the current distribution of astronomers, it would be logical to locate this facility in the eastern states (possibly in the Indiana/Ohio region) unless the system can take advantage of some existing computer network. Our experiment shows that astronomers do not differ significantly from the average user population in terms of session distribution. The load on the system during the day can therefore be predicted with some degree of accuracy.

2. Hardware.

The central computer will use a large storage device, such as a 2314 disk unit (possibly only half of it being used the first year) for the data base. Remote stations will use a terminal, an alphanumeric scope, and a medium-speed printer. No card or tape unit would be required as all large updates would be done centrally upon request. The nature of astronomical information is such that on-line update with full recovery is NOT a mandatory requirement of the system. The emphasis will be on interactive 'point-by-point' update and on fast interactive query. Under these conditions, the recovery capabilities of the current Stanford system have proved adequate and the same level of reliability can be achieved elsewhere without any special hardware. (See above, analysis of system failures). A local storage facility might be used in connection with


SESSION
DURATION
WEEKLY
AVERAGE IN
MINUTES
120
60
Week:
1
2
3
4
5
6
7
8


CPU
PERCENTAGE
OF DURATION
4
3
2
1
Week:
1
2
3
4
5
6
7
8


EDITING
PERCENTAGE
OF CPU TIME
30
20
10
Week:
1
2
3
4
5
6
7
8

the terminal, however, in the case of medium-size updates and catalogued interrogations. Such a device might be experimented with at one of the remote stations.

3. Software.

We feel our experiment has been most successful in showing that a text editor could be made to interface with a powerful retrieval language to achieve considerable flexibility in the implementation of a large, potentially unlimited data-base. The network presumably would rely on a combination similar to WYLBUR/DIRAC to achieve the same objective. Both systems would be re-entrant subprocessors in a time-sharing environment. (In our experiment, DIRAC was a FORTRAN prototype and was non-reentrant).

Conclusions

Under these assumptions, the major fraction of the operating cost would be spent on pure retrieval activities (this figure was 42% in our experiment). It is unlikely that performance improvement would decrease this overall cost, given the capacity of astronomical research to absorb immediately all the retrieval power it will be allowed to use. The same will be true of the text-editing CPU cost and of the storage costs. Communications and holding time are the two items that could be made more cost-effective in the full network, as compared with our experiment, although the bimodal distribution of holding times that we have observed may well prove typical in any update/query environemnt using scientific catalogues.

APPENDIX

AND BACKGROUND INFORMATION

"An Automatic Question-Answering System for Stellar Astronomy" By J.F.Vallee and J.A.Hynek. Reprinted from P.A.S.P. Vol.78, no.463, August 1966. Dearborn Observatory Contributions, no.37.

U.S.Naval Observatory Circular no.114. By Solomon Elvove, 16 January 1967.

U.S.Naval Observatory Circular no.104. By Victoria Meiller, 30 October 1964.

"The Henry Draper Catalogue on Magnetic Tape", note by D.Gottlieb and D.Hoffleit.

Correspondence between J.A.Hynek and P.Armer, 5 June 1970.

DEARBORN OBSERVATORY CONTRIBUTIONS

NO. 37

AN AUTOMATIC QUESTION-ANSWERING SYSTEM FOR STELLAR ASTRONOMY

JACQUES F. VALLEE AND J. ALLEN HYNEK

Reprinted from the *Publications of the Astronomical Society of the Pacific* (Vol. 78, No. 463, August 1966)

AN AUTOMATIC QUESTION-ANSWERING SYSTEM FOR STELLAR ASTRONOMY*

JACQUES F. VALLEE AND J. ALLEN HYNEK
Technological Institute and Lindheimer Astronomical Research Center, Northwestern University

Received February 14, 1966

Questions of a technical nature pertaining to the field of stellar evolution have been answered entirely by automatic means. In the research described, an English text submitted by the astronomer, consisting of a series of questions in machine-readable form, is scanned by a CDC 3400 data-processing machine and analyzed semantically. An information retrieval system automatically triggered by this analysis initiates a search through a star catalog and gives a numerical answer.

Introduction

In the last decades, the problem of storage and retrieval of large amounts of information has become a central one in many scientific disciplines, particularly those where heavy reliance is placed on statistical data. In stellar astronomy, for instance, the situation frequently arises in which catalogs need to be addressed at high speed. The Dearborn Observatory version (1965) of the Bright Star Catalogue, which contains space velocities both relative to the sun and to the local standard of rest, and which is available in machine-readable form, is a case in point. One might wish to use such a catalog to answer questions like the following:

How many binary systems have common proper motion components?

What is the proportion of spectroscopic binaries among all main-sequence stars which have only one visual component and whose total space velocity is less than 50 km/sec?

What is the percentage of giants later than F8 whose speed is below 70 km/sec?

What is the number of binaries in the Bright Star Catalogue whose primary is earlier than A3?

How many bright stars are double-line spectroscopic binaries?

Ordinarily, sorting schemes or separate programming would be necessary to extract this information (Weller 1964). In the research

* *Dearborn Observatory Contributions* No. 37.

to be discussed in this paper, we have designed a computer system that processes descriptive English questions such as the above and provides the corresponding answers at the average rate of eight questions per minute.

To achieve this result, we have first developed a compact code to express with only four characters the position of a star on the HR diagram, its spectral type, its multiplicity configuration, and its total space velocity. Next, we designed an information retrieval system capable of recognizing search strategies specified by "retrieval formulae," i.e. logical expressions which indicate without ambiguity the properties of the stars to be extracted from the Catalogue. Finally, a language processor was created which would accept natural language inputs and translate them into search strategies.

Algorithms for Converting the Characteristics of Stellar Systems into Codes

The *multiplicity code* of a stellar system is a standardized description of its configuration by an integer where each digit represents one component. In 1962, when the Dearborn version of the Bright Star Catalogue was converted to punched cards, a three-digit multiplicity code, where each system was considered as a potential visual triple, was defined (Table I).

Using this code, we can now express in logical terms the configuration of any system, as defined. In the following we shall call the *configuration index* of a stellar system a symbol expressing the

TABLE I

MULTIPLICITY CODE

	A First Component	B Second Component	C Third Component
0		None	None
1	Single star	Single star	Single star
2	Single-line sp. bin.	Single-line sp. bin.	Single-line sp. bin.
3	Double-line sp. bin.	Double-line sp. bin.	Double-line sp. bin.
4		Close companion	Close companion
5	Triple spectro. syst.	CPM, close pair	CPM, close pair
6	Composite spectrum	CPM, double-line sp. b.	CPM, double-line sp. b.
7		CPM, single-line sp. b.	CPM, single-line sp. b.
8	Astrometric binary	CPM, single star	CPM, single star
9	Single star, no RV	None	None

gravitational relationships observed between its members, by reference to an enumeration of all possible configurations. If (ABC) is the multiplicity code of a star, we can partition the set of values assigned to A, B, and C by defining the logical statements and the correspondences between system configurations and logical representations listed in Table II.

It will be noted that configurations 10 and 11 are rare, peculiar systems. In the following we simply denote all triple systems by the symbol T. This leaves only ten non-triple configurations, which can be thus denoted by a single digit between 0 and 9.

Now, let c be the configuration index defined above. Let h and s be symbols of the HR region of the star and its spectral range, respectively, as shown in Table III. Let V be the total space velocity of the system with respect to the centroid of mass of nearby stars* and expressed in kilometers per second.

We define the "velocity index" of a star, and denote it by the symbol v, as follows:

$v = 0$ if V is unknown
$v = n$ if $10(n-1) < V \leq 10(n)$
$v = 9$ if $V > 80$

TABLE II

DEFINITION OF SYSTEM CONFIGURATIONS

Logical Propositions	c	System	Logical Representation	No. of Systems
P1: A=1 or 9 ... "Primary is single"	0	0	P1 and Q1	6248
Q1: B=0 or 9 ... "No secondary"	1	•	non-P1 and Q1	1172
Q2: B=1 or 4 ... "Second. is single"	2	0-0	P1 and Q2	770
Q3: B=2 or 3 ... "Sec. is spec. bin."	3	•-0	non-P1 and Q2	140
Q4: B=8 ... "Sec. is CPM, single"	4	0-•	P1 and Q3	15
Q5: B=6 or 7 ... "Sec. is CPM, sp. bin"	5	•-•	non-P1 and Q3	16
R1: C=0 or 9 ... "System not triple"	6	0–0	P1 and Q4	334
Q6: B=5 ... "Sec. is visual bin."	7	•–0	non-P1 and Q4	83
Symbols:	8	0–•	P1 and Q5	7
0 is a single star	9	•–•	non-P1 and Q5	4
• is a spectroscopic binary	10	0–0-0	P1 and Q6	2
0-0 is a visual pair	11	•–0-0	non-P1 and Q6	1
0–0 is a CPM pair	12	TRIPLE	non-R1	141

* Space velocity components were computed by the Cracovian method (Przybylski 1962) for 5175 stellar systems.

TABLE III
HR INDEX AND SPECTRAL RANGE

HR Region	h	MK Class	Spectral range	Spectral classes	Center
Unknown	0		$s = a$	earlier than B2	—B0
			b	B3 – B7	B5
Dwarf	1	V	c	B8 – A2	A0
			d	A3 – A7	A5
Giant	2	III	e	A8 – F2	F0
			f	F3 – F7	F5
Supergiant	3	I, II	g	F8 – G2	G0
			h	G3 – G7	G5
Subgiant	4	IV	i	G8 – K2	K0
			j	K3 – K7	K5
Hertzsprung Gap	5		k	K8 – M2	M0
			l	later than M2	M5—
Subdwarf	6	VI			

The word of information $\sigma(S) = hscv$ is called the "signature" of a stellar system S.

The "signature" is of value in providing a compact, mnemonic summary of the outstanding characteristics of a stellar system. For instance, we see that 1c02 is the signature of a main-sequence star whose spectrum is in the range B8 to A2 and whose space velocity is between 10 and 20 km/sec. Similarly, 2k30 refers to a two-component system where the brightest star is a late K or early M giant and a spectroscopic binary, while the secondary is a single star, the space motion of the system being unknown.

Retrieval Formulae

In the context of this paper, any question regarding the characteristics of a stellar population can be of one of two types: one either asks for a number, namely the cardinality of a certain subset of the Catalogue, or the percentage or fraction of objects having certain properties among a certain sub-population whose cardinality may or may not have been previously determined.

We recognize the importance of this difference between the two types of questions by calling them "questions of type n," "questions of type p," respectively.

Example of a question of type n:

"What is the number of stellar systems whose space velocity is inferior to 20 km/sec?"

According to the velocity code we have defined (fourth digit of the "signature") we need the number of code-words which have as their last digit either a one (1) or a two (2). The answer is the sum of the cardinality of two subsets, and the retrieval formula will reflect this observation.

Example of a question of type p:

"What is the proportion of spectroscopic binaries among all main-sequence systems which have only one component and whose total space velocity is less than 50 km/sec?"

This question calls for two main operations. If we had to find the answer by manual sorting of cards, we would:

a) Decide to neglect the second digit of the signature, since the question does not involve the spectral types.

b) Reject from the Catalogue all codes which do not begin with a one (1) since we are only concerned with the main sequence (see Table III).

c) Reject all codes which do not have either a zero (0) or a one (1) as configuration index, since we want systems with one visual component only (Table II).

d) Select all codes whose last digit is 1, 2, 3, 4, or 5 (see definition of the velocity index).

After this series of operations, we would be left with a subset of the Catalogue, containing signatures of the following types:

1.01	1.02	1.03	1.04	1.05
1.11	1.12	1.13	1.14	1.15

We have used a period (.) to indicate that the second digit (spectral range) was to be ignored. In order to answer the question we now have to evaluate the cardinality of the ten subsets defined by these formulae, then eliminate all codes that do not have a one (1) as third digit, keeping only the spectroscopic binaries. The final answer is the ratio of the number of systems in the latter set to the preceding number.

The procedure that we follow in establishing this sorting scheme can be clarified by the following notation: We shall denote by $Y = (abcd)$ the cardinality of the subset of the Catalogue defined by considering only those stellar systems whose signature is *abcd*. The

two questions used as examples can now be restated in terms of this notation.

$Y = (...1) + (...2)$ answers the question "What is the number of stellar systems whose space velocity is inferior to 20 km/sec?" while

$$Y = (..1.)/((1.01) + (1.02) + (1.03) + (1.04) + (1.05) + (1.11) + (1.12) + (1.13) + (1.14) + (1.15))$$

defines the solution to the question we have just discussed in detail.

Such expressions are called *retrieval formulae*. It is natural to have a computer process these formulae and carry out the corresponding numerical manipulations, which are defined now without ambiguity. In this sense, retrieval formulae are a device for the mathematical representation of *search strategies*. An adequate formalism for the treatment of this problem is found in the theory of pushdown-store automata (to be published elsewhere by G. K. Krulee and Vallee).

The Linguistic Problem

In translating astronomical questions, expressed in ordinary English, into retrieval formulae, the difficulties we encounter are considerably smaller than those found in the general problem of machine translation. In our case, the object of language processing is to construct associations between descriptions of physical situations (expressed in the technical language of astronomy) and a machine configuration that extracts from the input statements formalized elements to be used in subsequent logical manipulations. The purpose of these manipulations is twofold:

1) To search for elements having certain properties, which should be extracted from the available records.
2) To perform certain operations on these subsets in order to answer a specific question.

Our solution to this problem is a program called ALTAIR (Automatic Logical Translation And Information Retrieval), where the input set of verbal statements has limitations which are ordinary in systems of this type. The operators *and*, *or*, *not* are prohibited, with the exception that the operator *and* may be used in ALTAIR within the list of attributes of a certain subset. Similarly, terms such as

TABLE IV

SAMPLE OF QUESTION-AND-ANSWER PERIOD

Problems and solutions	Cumulative Execution Time (minutes)
Q. COMPUTE THE NUMBER OF SYSTEMS HAVING THE HR REGION UNKNOWN. Y=(0...). A. THIS NUMBER IS 2298	0.00
Q WHAT IS THE PROPORTION OF SYSTEMS HAVING STARS. A. 8933 STARS OUT OF 8933, OR 100.0 PERCENT	0.04
Q. REQUEST ALL STARS REDDER THAN F8 AMONG MAIN SEQUENCE SYSTEMS. A. QUESTION NOT COMPATIBLE WITH CURRENT CODING SYSTEM AND/OR SEMANTICALLY INCORRECT	0.04
Q. REQUEST ALL MAIN SEQUENCE SYSTEMS REDDER THAN F8. Y=((1G..)+(1H..)+(1I..)+(1J..)+(1K..)+(1L..)) A. THIS NUMBER IS 331	0.04
Q. REQUEST ALL STARS BETWEEN F3 AND F7 HAVING SPEED ABOVE FORTY. Y=((.F.5)+(.F.6)+(.F.7)+(.F.8)+(.F.9)) A. THIS NUMBER IS 80	0.25
Q. WHAT IS THE PROPORTION OF STARS HAVING A SPEED ABOVE 40 KM/SEC AMONG THE SYSTEMS WHOSE SPECTRAL TYPE IS BETWEEN F3 AND F7. Y=((...5)+(...6)+(...7)+(...8)+(...9))/(.F..) A. 80 STARS OUT OF 534, OR 15.0 PERCENT	0.43
Q. COMPUTE THE NUMBER OF STARS OF MK CLASS III WHOSE SPECTRAL TYPE FALLS BETWEEN F3 AND K2. Y=((2F..)+(2G..)+(2H..)+(2I..)) A. THIS NUMBER IS 1463	0.49
Q. WHAT IS THE PROPORTION OF SUBGIANTS AMONG THOSE STARS WHICH ARE OF SPECTRAL TYPE EARLIER THAN F7. Y=(4...)/((.A..)+(.B..)+(.C..)+(.D..)+(.E..)+(.F..))	0.63

A. 333 STARS OUT OF 4812, OR 6.9 PERCENT

Q. HOW MANY SYSTEMS OF MK CLASS IV ARE INCLUDED IN THE BRIGHT STAR CATALOGUE. 0.88

A. QUESTION NOT COMPATIBLE WITH CURRENT CODING SYSTEM AND/OR SEMANTICALLY INCORRECT

Q. HOW MANY MULTIPLE STARS ARE DOUBLE. 0.88

Y=((..2.)+(..3.)+(..4.)+(..5.)+(..6.)+(..7.)+(..8.)+(..9.)+(..T.))

A. THIS NUMBER IS 1513

Q. HOW MANY STELLAR SYSTEMS HAVE PRIMARIES. 1.22

Y=(....)

A. THIS NUMBER IS 8933

Q. WHAT IS THE PROPORTION OF SYSTEMS HAVING A VELOCITY IN EXCESS OF 80 KM/SEC AMONG ALL STARS WHICH HAVE A COMMON PROPER MOTION COMPANION. 1.29

Y=(...9)/((..6.)+(..7.)+(..8.)+(..9.))

A. 8 STARS OUT OF 428, OR 1.9 PERCENT

Q. FIND THE PERCENTAGE OF VISUAL PAIRS AMONG ALL STELLAR SYSTEMS WHERE THE PRIMARY IS REDDER THAN F3 AND WHOSE VELOCITY IS BELOW 20 KM/SEC. 1.44

A. RETRIEVAL FORMULA LENGTH EXCEEDED

most or *highest* are prohibited. Such restrictions, as pointed out by the developers of BASEBALL are minor ones and can easily be removed at a later stage of refinement of the technique. In ALTAIR, however, the input sentence is not restricted to the form of a single-clause question since it was designed specifically to process questions of type p.

We give as illustration in Table IV a series of questions of the type considered in this article, as well as the answers given by ALTAIR. The language processor is capable of generating retrieval formulae from English statements at the average rate of 250 formulae per minute. The searching time through the Catalogue, however, is longer, and the resulting performance of the program is about eight answers per minute. We have purposely included some ques-

tions that were irrelevant to the problem context and were accordingly rejected by the program. Note that the questions were addressed to the Dearborn revised Bright Star Catalogue and were verified by reference to conventional sorting procedures.

As a linguistic processor, ALTAIR is still, of course, at an experimental stage. Implementation of the system requires access to a large computer, and our current tests are run on a CDC 3400. For most practical purposes, however, those wishing to take advantage of the flexibility of the scheme of *retrieval formulae* could very easily implement our method on practically any computer available on the market, by addressing directly the searching automaton. This would imply abandoning the idea of direct access in natural language and assumes that user has been trained to write retrieval formulae. The same scheme could readily be applied to any astronomical catalog by defining an appropriate system of codes similar to our system of "signatures."

Conclusion

We have tried to demonstrate the feasibility of an inquiry system driven by questions in ordinary English. It gives the astronomer direct access to Catalogue data and frees him from the necessity of wording his questions in the language of a particular data-processing system. Because of the large computer required, such a system will find its greatest efficiency in a time-sharing environment.

This research is supported as part of an interdisciplinary project conducted jointly by the astronomy department and the department of industrial engineering and management science at Northwestern University. The writers are indebted to Professor Gilbert K. Krulee for his encouragement and valuable guidance in designing this system.

REFERENCES

Dearborn Observatory Version of the Yale Bright Star Catalogue (1965) unpublished, available at the Observatory Library, Evanston.
Przybylski, A. 1962, *Acta Astronomica* 12, No. 4; *Mount Stromlo Obs. Reprint* No. 68.
Weller, W. J. 1964, *Pub. A.S.P.* 76, 152.

UNITED STATES NAVAL OBSERVATORY

CIRCULAR NO. 114

U. S. Naval Observatory, Washington, D. C. 20390 January 16, 1967

ASTRONOMICAL DATA IN MACHINE READABLE FORM

by

Solomon Elvove

This *Circular* supersedes *Circular* 111, 14 January 1966. It contains a revised list, with occasional changes in item number, of the astronomical data available in machine readable form from the U. S. Naval Observatory. Duplicates of these data, on punched cards or magnetic tape, will be made in response to requests from astronomical observatories or other scientific laboratories. Inquiries relative to procurement of copies should be directed to the Superintendent, U. S. Naval Observatory, Washington, D. C. 20390.

(1) *The American Ephemeris and Nautical Almanac*

(a) Through 1980: Daily values at 0^h E. T. of: apparent right ascension and declination, radius vector, semidiameter. Equation of time through 1967; ephemeris transit beginning 1968 (365 cards per year).
(b) 1968–70: Daily values at 0^h E. T. of: longitude (mean equinox of year), reduction to apparent longitude, latitude (ecliptic of 1950.0, year, and date), horizontal parallax, precession in longitude, nutation in longitude and obliquity of ecliptic (365 cards per year).
(c) Ephemeris for physical observations, through 1970 (365 cards per year).

(2) *Astronomical Papers*, Vol. XIV

(a) left hand pages: spherical and rectangular coordinates of the Earth–Moon barycenter referred to 1950.0, 1800–2000 (23,000 cards); tapes available in 4, 10, and 20 day intervals.
(b) right hand pages: geocentric spherical (referred to date) and rectangular (referred to 1950.0) coordinates of the Sun, 1800–2000 (19,000 cards).

(3) Geocentric equatorial rectangular coordinates in units of Earth's radius

(a) referred to mean equinox of nearest beginning of year, every 2 days, 1957 December 16 through 1971 January 13 (2400 cards),
(b) referred to 1950.0, every 2 days, 1957 December 15 through 1971 January 12 (2400 cards),
(c) referred to 1963.0, daily, 1959 January 1 through 1971 January 12 (4800 cards).

(4) Apparent longitude, referred to true equinox of date

(a) for every 10 days, 100 B. C. – A. D. 1800 (7,000 cards),
(b) daily, 1850 – 2000. Magnetic tape only.

(5) *Publications of the U. S. Naval Observatory*

Positions of the sun, moon and planets observed with the Six-Inch Transit Circle

(a) Vol. XVI
 (1) Part I, 1925 – 1941 (7,615 cards).
 (2) Part III, 1941– 1949 (2,565 cards).
(b) Vol. XIX
 (1) Part I, 1949 – 1956 (3,043 cards).

(6) *U. S. Naval Observatory Circulars* Nos. 103, 105 and 108

Positions of the sun, moon and planets observed with the Six-Inch Transit Circle between 1956 and 24 December 1964 (4,220 cards).

Also see Miscellaneous: *The Nautical Almanac*, right hand pages and *The Air Almanac*.

Moon

(1) *The American Ephemeris and Nautical Almanac* and *Improved Lunar Ephemeris*. 1952–1959

(a) Apparent longitude and latitude, semidiameter, horizontal parallax, ephemeris transit, for every half day through 1971. Includes nutation and aberration (730 cards per year).

(b) Apparent right ascension and declination, horizontal parallax, hourly through 1971 (8,800 cards per year). Also available on magnetic tape, blocked 24 per record.

(c) Ephemeris for physical observations, through 1970 (365 cards per year).

(2) *The Nautical Almanac*, right hand pages: Greenwich hour angle, apparent declination, horizontal parallax, sidereal time, hourly through 1968 (8,800 cards per year).

(3) Geocentric equatorial rectangular coordinates in units of Earth's radius for every half day, 1952 January 1 through 1971 July 4

(a) referred to mean equinox of nearest beginning of year, as published in *U. S. Naval Observatory Circular* No. 91, *Rectangular Coordinates of the Moon*, 1952–1971 (730 cards per year),

(b) referred to 1950.0 (730 cards per year),

(c) referred to 1963.0 (730 cards per year).

(4) Longitude, latitude, horizontal parallax, referred to mean equinox of date for every half day through 1971. Includes aberration except for the term, $+0''.018 \cos (l - 2D) + 0''.007 \cos 2D$ of the aberration (730 cards per year).

(5) Longitude, latitude, horizontal parallax, semidiameter, obliquity, nutation in longitude, nutation in obliquity, referred to mean equinox of date for every half day, 1850–1999, inclusive. Includes aberration. On magnetic tape only, six reels: (a) 1850–1874, (b) 1875–1899, (c) 1900–1924, (d) 1925–1949, (e) 1950–1974, (f) 1975–1999. In preparation.

(6) Apparent longitude and latitude, referred to true equinox of date, daily, 300 B. C. – A. D. 2074. Magnetic tape only.

(7) *Publications of the U. S. Naval Observatory*

Positions of the sun, moon and planets observed with the Six-Inch Transit Circle

(a) Vol. XVI
- (1) Part I, 1925 – 1941 (7,615 cards).
- (2) Part III, 1941 – 1949 (2,565 cards).

(b) Vol. XIX
- (1) Part I, 1949 – 1956 (3,043 cards).

(8) *U. S. Naval Observatory Circulars* Nos. 103, 105 and 108

Positions of the sun, moon and planets observed with the Six-Inch Transit Circle between 1956 and 24 December 1964 (4,220 cards).

Also see Miscellaneous: *The Nautical Almanac*, right hand pages and *The Air Almanac*.

Planets

(1) *The American Ephemeris and Nautical Almanac*

(a) Apparent right ascension and declination, semidiameter, horizontal parallax, true distance from Earth, ephemeris transit, daily through 1967. All planets except Pluto (365 cards per year per planet).
(b) Apparent right ascension and declination, true distance from Earth and ephemeris transit, daily, 1968–1980, inclusive. All planets except Pluto (365 cards per year per planet).
(c) Astrometric right ascension and declination, horizontal parallax, true distance from Earth, ephemeris transit of Pluto, for every 4 days through 1971 (92 cards per year).
(d) Heliocentric longitude and latitude, radius vector, orbital longitude, daily motion of orbital longitude of Jupiter and Saturn, referred to mean equinox and ecliptic of date, for every 10 days, 1960 May 7 through 1982 January 31 (801 cards per planet).
(e) Heliocentric longitude and latitude, radius vector, orbital longitude, daily motion of orbital longitude of Uranus, Neptune and Pluto, referred to mean equinox and ecliptic of date, for every 40 days, 1959 December 18 through 1982 (205 cards per planet).
(f) Ephemeris for physical observations of Mercury, Venus, Mars, Jupiter and Saturn through 1970.

(2) *Astronomical Papers*, Vol. XII, *Coordinates of the Five Outer Planets*, 1653–2060

(a) 40–day values (3,721 cards per planet).
(b) 10–day values, subtabulated from 40–day coordinates, by date and planet. Magnetic tape only.
(c) 40–day values, by date and planet. Magnetic tape only.

(3) *Astronomical Papers*, Vol. XV, Part III, *Heliocentric Coordinates of Venus*, 1800–2000

(a) 4–day values (41,000 cards).
(b) 4–day values, blocked 10 to a record. Magnetic tape only.
(c) 10–day values, unblocked. Magnetic tape only.
(d) 20–day values, blocked 10 to a record. Magnetic tape only.

(4) *U. S. Naval Observatory Circulars* Nos. 90 and 95

Provisional ephemeris of Mars based on Clemence's theory. Heliocentric equatorial rectangular coordinates, referred to mean equinox and equator of 1950.0, for every 4 days (a) 1800–1950 (tape only), (b) 1950–2000 (2,350 cards).

(5) *U. S. Naval Observatory Circular* No. 96

Geocentric distance and velocity of Venus referred to 1950.0, daily, 1960–1970, inclusive (365 cards per year).

(6) *U. S. Naval Observatory Circular* No. 98

Physical ephemeris of Mars based on a position of the north pole of $\alpha = 316.^\circ55 + 0.^\circ006533\,(t - 1905.0)$, $\delta = +52.^\circ85 + 0.^\circ003542\,(t - 1905.0)$, computed at a two day interval for sixty days centered at each opposition from 1877–1967 (5,246 cards).

(7) *Planetary Co–ordinates*, 1960–1980

(a) Heliocentric equatorial rectangular coordinates of Mercury, referred to 1950.0, for every 5 days (365 cards).
(b) Longitude, latitude, radius vector and equatorial rectangular coordinates of Venus, Earth and Mars, referred to 1950.0, for every 10 days (769 cards per planet).

Planets

(8) Heliocentric equatorial rectangular coordinates and first derivatives of Venus and Earth–Moon barycenter, referred to 1950.0 daily, 1960–1970, inclusive

(a) Coordinates (365 cards per year).
(b) Derivatives (365 cards per year).

(9) Longitude, latitude and radius vector of Mercury, Venus and Mars, daily through 1980 (365 cards per year per planet).

(10) (a) Heliocentric equatorial rectangular coordinates of all planets except Mercury, referred to 1950.0, by date and planet; for every 10 days 1800 January 4.5 through 2000 December 2.0. Venus, same as 3(c); Earth–Moon barycenter, same as Sun, 2(a); Mars, unpublished Herget integration; outer planets, same as 2(b). Magnetic tape only.
(b) Same as 10(a) but in reverse chronological order. Magnetic tape only.

(11) *Publications of the U. S. Naval Observatory*

Positions of the sun, moon and planets observed with the Six–Inch Transit Circle

(a) Vol. XVI
  (1) Part I, 1925 – 1941 (7,615 cards).
  (2) Part III, 1941 – 1949 (2,565 cards).
(b) Vol. XIX
  (1) Part I, 1949 – 1956 (3,043 cards).

(12) *U. S. Naval Observatory Circulars* Nos. 103, 105 and 108

Positions of the sun, moon and planets observed with the Six–Inch Transit Circle between 1956 and 24 December 1964 (4,220 cards).

(13) *U. S. Naval Observatory Circular* No. 106

Rectangular co-ordinates of Mercury, 1800–2000. Magnetic tape only.

Also see Miscellaneous: *The Nautical Almanac*, left hand pages and *The Air Almanac*.

Minor Planets

(1) *The American Ephemeris and Nautical Almanac*

Astrometric right ascension and declination, true distance from Earth, ephemeris transit of Ceres, Pallas, Juno, Vesta, daily through 1971 (365 cards per year per planet).

(2) *Astronomical Papers*, Vol. XI, Part IV and Vol. XVI, Part III

Heliocentric equatorial rectangular coordinates of Ceres, Pallas, Juno and Vesta, referred to 1950.0 for every 20 days, 1920–1960 (770 cards per planet); for every 10 days, 1960–1980 (770 cards per planet).

Star Catalogues

Available

(1) *Publications of the United States Naval Observatory*

- (a) Vol. VIII (1,201 cards).
- (b) Vol. IX
  - (1) Part I (4,526 cards).
  - (2) Part IV (196 cards).
- (c) Vol. X, Part I (368 cards).
- (d) Vol. XI
  - (1) Catalogue of 215 Stars (215 cards).
  - (2) Catalogue of 2,499 Stars (2,499 cards).
- (e) Vol. XIII
  - (1) (10,571 cards).
  - (2) Appendix II (818 cards).
- (f) Vol. XIV
  - (1) Part I (3,520 cards).
  - (2) Part II (1,248 cards).
  - (3) Part III (2,094 cards).
- (g) Vol. XV, Part V (5,446 cards).
- (h) Vol. XVI
  - (1) Part I
    - (a) Stars for 1925.0 (2,383 cards).
    - (b) Stars for 1950.0 (1,536 cards).
  - (2) Part III (5,216 cards).
- (i) Vol. XIX, Part I (5,965 cards).

(2) *Astronomical Papers*, Vol. XIII, Part III, N30 (5,268 cards).

(3) Boss, *Albany General Catalogue* (33,342 cards).

(4) Cross References:

- (a) Boss Catalogue star number, Henry Draper star number (3,335 cards).
- (b) Boss Catalogue star number, Henry Draper star number, FK3 star number, Durchmusterung number (33,342 cards).
- (c) Same as 4 (b) but in order of Henry Draper number. Magnetic tape only.

(5) Yale Observatory Catalogues

Declinations	Cards	Declinations	Cards
$+55^\circ$ to $+60^\circ$	8,164	-2° to $+1^\circ$	5,583
$+50$ to $+55$	8,380	-2 to -6	8,108
		-6 to -10	8,248
$+25$ to $+30$	10,358	-10 to -14	8,101
$+20$ to $+25$	8,703	-14 to -18	8,563
$+15$ to $+20$	9,092	-18 to -20	4,553
$+10$ to $+15$	8,967	-20 to -22	4,292
$+9$ to $+10$	1,906	-22 to -27	15,110
$+5$ to $+9$	9,060	-27 to -30	9,455
$+1$ to $+5$	7,996		

Star Catalogues

Available

(6) Yale Catalogue of Bright Stars (Third Revised Edition, 1964). Magnetic tape only.

(7) Wilson, *General Catalogue of Stellar Radial Velocities*, 1953 (15,106 cards).

(8) FK3

(a) Right ascension (1,591 cards).
(b) Declination (1,591 cards).

(9) FK4 (1,535 cards).

(10) AGK2

Declinations ° °	Cards	Declinations ° °	Cards
+ 70 to + 89	11,801	+ 25 to + 29	14,367
+ 60 to + 69	13,064	+ 20 to + 24	12,992
+ 55 to + 59	8,535	+ 15 to + 19	12,857
+ 50 to + 54	9,168	+ 10 to + 14	14,330
+ 45 to + 49	10,915	+ 5 to + 9	16,770
+ 40 to + 44	11,806	0 to + 4	15,094
+ 35 to + 39	12,337	0 to − 2.5	7,157
+ 30 to + 34	12,395		

(11) *Geschichte des Fixsternhimmels*. δ 0° to + 90°, for 1875 (tape only).

(12) Kustner, Catalogue of 10,663 Stars. Bonn, Vol. 10, δ 0° to +51°(10,663 cards).

(13) Berlin, Vol. 55, AGK2A (13,747 cards).

(14) *Cape Zone Catalogue for* 1900

(a) δ −40° to −52°(20,843 cards).
(b) Also precessed to 1950.0 (20,843 cards).

(15) *Cape Photographic Catalogue for* 1950

Declinations ° °	Cards	Declinations ° °	Cards
− 30 to − 35	12,846	− 56 to − 60	7,657
− 35 to − 40	12,115	− 60 to − 64	7,053
− 52 to − 56	9,215		

(16) First Cape Catalogue for 1925 (4,569 cards).

(17) Second Cape Catalogue for 1925 (11,667 cards).

(18) Third Cape Catalogue for 1925. δ 0° to −30°(6,597 cards).

(19) First Cape Catalogue for 1950 (5,070 cards).

(20) Semirot, Bordeaux Meridian Circle Observations, δ +11° to +18°(2,024 cards).

(21) Fayet, Meridian Circle at Nice, δ +5° to +15°(964 cards).

Star Catalogues

In Preparation

(22) Second Washington Catalogue for 1875 (5,429 cards).

(23) Yale, *General Catalogue of Trigonometric Stellar Parallaxes*, 1952 (6,080 cards).

(24) Publications of Pulkova Observatory, Vol. LXII

(a) Left hand pages (587 cards).
(b) Right hand pages (587 cards).

(25) Kukarkin, General Catalogue of Variable Stars

(a) Second edition (14,708 cards).
(b) First supplement to the second edition:
(1) (1,647 cards).
(2) (796 cards).

(26) *Albany Catalogue* for 1910.0 (20,811 cards).

Appendix (1,125 cards).

(27) Boss, *Preliminary General Catalogue of 6,188 Stars* for 1900 (6,188 cards).

(28) *Cape Photographic Catalogue for* 1950

Declinations	Cards
−76° to −80°	2,934
−72 to −76	3,764
−68 to −72	4,503
−64 to −68	5,398

(29) Boss, *San Luis Catalogue for* 1910 (15,333 cards).

(30) Cordoba Fundamental General Catalogue for 1950.0 (761 cards).

(31) *Astronomical Papers*

(a) Vol. VIII

(1) Part II (1,596 cards).
(2) Part III (1,607 cards).

(b) Vol. X (1,504 cards).

(32) *Cape Fundamental Catalogue for* 1900 (1,293 cards).

(33) *Second Cape Fundamental Catalogue for* 1900 (1,846 cards).

(34) *Cape General Catalogue for* 1900 (4,464 cards).

(35) *Cape Catalogue of* 1680 *Stars for* 1900 (1,680 cards).

Miscellaneous

(1) Sidereal Time, through 1980 (366 cards per year).

(2) Equation of Time for 0^h U. T., 1965–1970 (62 cards per year).

(3) Nutation in longitude and obliquity, 1949 March 25 through 2009 June 18 (365 cards per year).

(4) Besselian and Independent Day Numbers, 1960–1971

(a) referred to equinox of 1950.0 (365 cards per year).
(b) referred to nearest beginning of a year (365 cards per year).

(5) *The Nautical Almanac*, 1964 through 1969

(a) Left hand pages: Greenwich hour angle and declination of Venus, Mars, Jupiter and Saturn, hourly; sidereal hour angle and declination of 57 stars for every third day (9,000 cards per year).
(b) Right hand pages: Greenwich hour angle and declination of Sun and Moon, hourly; sunrise–sunset, civil and nautical twilight, moonrise–moonset (9,000 cards per year).

(6) *The Air Almanac*, 1964–1968 (53,000 cards per year).

Coordinates

Geocentric equatorial spherical

- Sun (1a) (5a) (5b) (6)
- Moon (1b) (7a) (7b) (8)
- Planets (1a) (1b) (1c) (11a) (11b) (12)
- Minor Planets (1)
- *The Nautical Almanac* [Miscellaneous (4a) (4b)]
- *The Air Almanac* [Miscellaneous (5)]

Geocentric equatorial rectangular

- Sun (2b) (3a) (3b) (3c)
- Moon (3a) (3b) (3c)

Geocentric ecliptic spherical

- Sun (1b) (2b)
- Moon (1a) (4) (5)

Heliocentric equatorial rectangular

- Earth–Moon barycenter [Sun (2a)]
- Planets (2a) (2b) (2c) (3b) (3c) (3d) (4) (7a) (7b) (8a) (8b) (10a) (10b) (13)

Heliocentric ecliptic spherical

- Earth–Moon barycenter [Sun (2a)]
- Planets (1d) (1e) (3a) (7b) (9)

UNITED STATES NAVAL OBSERVATORY

CIRCULAR NO. 104

U. S. Naval Observatory, Washington, D. C. 20390 October 30, 1964

DURCHMUSTERUNG AND HENRY DRAPER NUMBERS OF ALBANY GENERAL CATALOG STARS

by

Victoria Meiller

This correlation of Henry Draper and Bonner, Cape, or Cordoba Durchmusterung Star Numbers with the stars of the *Albany General Catalog* has been produced as part of the cooperative Star Catalog Project of the Yale University Observatory and the Nautical Almanac Office, U. S. Naval Observatory. The Henry Draper Numbers were provided as a result of information furnished by the Astronomischen Rechen-Institut.

This correlation is arranged in the order of the Albany General Catalog Number. A similar correlation in the order of the Henry Draper Number is available on magnetic tape and listings will be furnished upon request. A correlation of Albany General Catalog and Bonner, Cape, or Cordoba Durchmusterung Star Numbers with FK3 Star Numbers is given in *U. S. Naval Observatory Circular* No. 93.

The Henry Draper Catalogue on Magnetic Tape

AN extensive program to punch the *Henry Draper Catalogue* (A. J. Cannon and E. C. Pickering 1918–24, *Ann. Astron. Obs. Harvard Coll.* 91–99) and its *Extension* (A. J. Cannon 1925–36, *Ann. Astron. Obs. Harvard Coll.* 100) on cards, was instituted by D. Hoffleit at Yale University in 1962. The project was completed in 1966, and the cards read onto tape (unblocked, low density). During the preparation of the punched cards, some 400 errors in the printed catalogues were corrected. The majority of these had been reported by Mrs. Margaret Walton Mayall (at one time Miss Cannon's chief assistant); about 100 others were detected in the course of the key-punching and subsequent usage of the cards. All of this work was accomplished under U. S. Naval Observatory contract NONR 609 (41) for a joint Yale–U. S. Naval Observatory project for putting major stellar catalogues on punch cards.

A copy of the catalogue, at this point occupying nine 2400-ft tapes, was sent to the University of Maryland. Here, the nine tapes were consolidated onto one by blocking 126 stars together and using high density. In addition, a small number of sequencing errors were corrected. The final tape now contains eight quantities per star: The HD number, the BD number, the right ascension and declination (1900.0) in degrees, the photometric and photographic magnitude, a numerical code for the spectral type, and a word of data indicating pecularities.

This tape is in order of HD number (thus increasing right ascension); another tape is available where the catalogue has been rearranged into declination zones. There are 18 such zones, each 12° wide, allowing for a 1° overlap between zones. Within the declination zone, the catalogue is in order of right ascension.

A third tape contains similar data for the first volume of the *Henry Draper Extension*, in order to right ascension. (The second volume of the Extension, *Harvard Ann.* 112, was not punched because the data are available only on star charts.)

All three tapes are readable on IBM 7094's using IOCS, and on other machines using simple assembly language programs. For persons not having adequate computing or programming facilities, we are willing to generate printouts of the HD for specialized groups of stars.

Further details, and the tapes themselves, are available at a cost to cover the expense of duplication from the Astronomy Program, University of Maryland, College Park, Maryland 20742.

D. GOTTLEIB, *University of Maryland*
D. HOFFLEIT, *Yale University Observatory*

J. Vallee

NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60201

DEPARTMENT OF ASTRONOMY

LINDHEIMER ASTRONOMICAL RESEARCH CENTER

5 June 1970

Dr. Paul Armer, Director
Computation Center
Stanford University
Palo Alto, California 94305

Dear Dr. Armer:

Now that the Stanford-Northwestern joint experiment in the testing of WYLBUR/DIRAC is over, I want to thank you and your colleagues, on behalf of the Department of Astronomy and myself personally, for making this effort possible. We certainly have profited from it; it was especially important for us to have had this experience in the use of a sophisticated time sharing system. We hope, too, that our use of the Stanford systems served to provide you with some idea of the ways in which scientists will work with such systems in the future, even though all patterns of use surely cannot be foreseen clearly.

I think that the field of astronomy is a particularly good one in which to have performed this experiment, since in astronomical data gathering there are such a vast number of data points. Every star in the sky is the potential source of literally hundreds of data points, and the old-fashioned methods of compiling astronomical catalogues of the many types of data are literally beginning to fall down under their own weight. The present system is cumbersome and constitutes a growing obstacle to further progress. Clearly the astronomical fraternity would do well to adopt a system similar to that which we have been testing.

It may be of interest for you to know that at the forthcoming meetings of the American Astronomical Society in Boulder, Colorado, I will be meeting with the American section of the International Astronomical Union Committee on Astronomical Data Processing and will present, with Lloyd Wackerling, a synopsis of the Stanford-Northwestern Experiment. Later this summer a similar synopsis will be presented to the International Astronomical Union at its Triennial Meeting in Brighton, England. Certainly the experience we have gained will be invaluable in discussing with interested colleagues the best means of presenting a proposal to supporting agencies for the continuation of this type of work. We recognize that the current economic stringencies may entail some delays in putting the ideas of Dr. Vallee and those of my colleagues into immediate effect, but despite these restrictions, we continue to receive encouraging remarks from other colleagues on the possibility of their, at least, moral support for these new concepts of astronomical data processing.

I will keep you informed through Dr. Vallee of the results of our talks in Boulder and later on at Brighton. At both places we have asked interested astronomers to meet with us to discuss the proposed system.

Once again, may I extend my personal thanks and those of my colleagues for your kindness and cooperation in having made our joint experiment a success. I was actually surprised personally by the number of our staff and students who actively used the terminal and profited greatly therefrom.

Sincerely yours,

J. Allen Hynek

J. Allen Hynek
Director

JAH:wpr